Old Dominion University

ODU Digital Commons

Psychology Theses & Dissertations   Psychology

Fall 12-2019

Examining the Effect of Interruptions at Different Breakpoints and Frequencies Within a Task

Sarah Anastazia Powers
Old Dominion University, sarahpowers623@gmail.com

Recommended Citation

EXAMINING THE EFFECT OF INTERRUPTIONS AT DIFFERENT

BREAKPOINTS AND FREQUENCIES WITHIN A TASK

by

Sarah Anastazia Powers
B.S. May 2016, Iowa State University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2019

Approved by:

Mark W. Scerbo (Director)

Yusuke Yamani (Member)

Ivan Ash (Member)

ABSTRACT

EXAMINING THE EFFECTS OF INTERRUPTIONS AT DIFFERENT
BREAKPOINTS AND FREQUENCIES WITHIN A TASK

Sarah Anastazia Powers
Old Dominion University, 2019
Director: Dr. Mark W. Scerbo

Research on the effects of interruptions shows that they can be either costly or beneficial depending on which aspects of an interruption are manipulated. One important aspect that contributes to these conflicting results concerns when an interruption occurs. The present study explored how event segmentation theory (EST) can be used to determine optimal moments for an interruption relying on hierarchical task analysis (HTA) to identify coarse and fine event boundaries. Utilizing a 2 X 3 mixed design, undergraduate students completed a trip planning task divided into three task trials. The within-subjects factor was interruption timing, which had three levels: none, coarse breakpoints, and fine breakpoints. The between-subjects factor was interruption frequency, which had two levels: one and three. According to memory for goals theory (MFG), a task representation at a fine breakpoint is large and thus an interruption occurring at this breakpoint increases memory demand and results in performance decrements when compared to an interruption at a coarse breakpoint. In line with this theory, it was hypothesized that interruptions would be more disruptive at fine vs. coarse breakpoints and that as the frequency of interruptions increased, so would the degree of disruption. Last, it was expected that the effects of high frequency interruptions would be more pronounced at fine vs. coarse breakpoints. The dependent measures included resumption lag, task completion time, number of errors, mental workload, and frustration. The findings provided partial support for

these predictions. Consistent with MFG theory and EST, participants took longer to resume the primary task and reported higher mental workload and frustration when interruptions occurred at fine breakpoints. Conversely, the effects of interruptions at coarse breakpoints were similar to completing the task without interruption. However, interruption frequency had no effect on performance. In general, these results suggest that the disruptiveness of a single or even a few interruptions is tied to the point within the task hierarchy where it occurs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The need to study interruptions is important as research brings awareness to the negative impact interruptions have on performance (Gillie & Broadbent, 1989; Trafton & Monk, 2007). In a review of 63 experimental studies, Li and colleagues (2011) describe some consequences of interruptions such as negative effects on task completion time, work strategy, decision-making, and error rate. In fact, an individual is twice as likely to make an error and take longer to complete the task when interrupted than when not interrupted (Flynn et al., 1999). Similarly, Iqbal and Bailey (2005) found that individuals who are interrupted experience increased frustration, annoyance, and workload. In fact, interruptions can be so disruptive that common office staff report spending more than a quarter of their workday attending to and recovering from interruptions (Spira & Feintuch, 2005). In high stakes environments such as healthcare, the effects of interruptions are even more alarming and can result in missing crucial patient information (Grundgeiger & Sanderson, 2009) or dispensing incorrect medication (Flynn et al., 1999).

Interruptions are increasingly prevalent across domains such as healthcare, aviation, transportation, energy, marketing, military, and others (Grundgeiger, Liu, Sanderson, Jenkins, & Leane, 2008; Morgan, Patrick, Waldron, King, & Patrick, 2009; Trafton & Monk, 2007). As a result, interruptions are often examined through different theoretical lenses, making it difficult to develop an overarching definition. In an attempt to fuse competing definitions, a literature review conducted by Brixey et al. (2007) examined 120 articles focused on interruptions from numerous disciplines. After a thorough analysis, they described an interruption by several core

attributes: discontinuity; externally initiated; intrusion of secondary, unplanned, and unexpected task; and situated within a context.

Discontinuity refers to a break in continuous task performance by the interruption. This discontinuity manifests in two ways, an interruption lag and a resumption lag. When an individual is preparing to switch from the primary task to the interrupting task, preparatory time is needed and constitutes the interruption lag. Discontinuity also occurs when an individual transitions from completing the interrupting task back to the suspended primary task. The time taken to resume the primary task is known as the resumption lag. Interruptions are externally initiated, meaning they are generated via another individual, medium, or device. Intrusion of a secondary, unplanned, and unexpected task represents the idea that an interruption can occur at any time and requires that another task is addressed. Finally, interruptions occur within a context such as a location or setting.

These core attributes not only define what it means for something to be considered an interruption, but they additionally serve as exclusionary criteria for what is not considered an interruption. For example, whereas interruptions are initiated externally, distractions can be both external or internal in nature. Externally initiated distractions stem from events in the environment (Mark, Gonzalez, & Harris, 2005), such as overhearing a loud conversation while studying (Robison & Unsworth, 2015). Internally initiated distractions, such as daydreaming, are competing thoughts unrelated to the primary task (Brixy et al., 2007). Also, unlike an interruption, a distraction does not involve discontinuity in task performance (Flynn et al., 1999). Although both interruptions and distractions occur concurrently with the primary task, an interruption requires suspension of the task so that it can be addressed immediately. By contrast, distractions are considered irrelevant stimuli that can be disregarded or deferred at the recipient's

discretion while still maintaining activation of encoded information (Clapp & Gazzaley, 2012; Clapp, Rubens, & Gazzaley, 2009). Not all interruptions are irrelevant, some may be necessary for the primary task.

Despite the consequences associated with interruptions, the results of several studies suggest that interruptions are not only unavoidable but can even be beneficial under the right circumstances. Many professional workplaces depend on interruptions to ensure success (Healey, Primus, & Koutantji, 2007). For example, if a pilot is approaching a runway and the landing gear has malfunctioned, it becomes critical and appropriate to interrupt the pilot with this information (Hayes & Miller, 2010). Circumstances such as these support the idea that interruptions may be used as an effective tool when implemented properly. O'Conaill and Frohlich (1995) were some of the first researchers to explore the possibility that interruptions could be used advantageously, reporting that in 64% of the time, an interruption was helpful to the recipient. Some benefits include: providing important information, temporary relief from a task, or redirecting attention to a more important task (Spira & Feintuch, 2005). Useful interruptions can come in the form of notifications on cell phones (Bailey & Iqbal, 2008), computers (Bailey & Konstan, 2006), and even in vehicles (Parasuraman, Cosenzo, & De Visser, 2009). Although notifications are perceived as helpful and can convey crucial information, they still demand immediate attention and can occur at inopportune times (Chen, Hart, & Vertegaal, 2007). Due to characteristics like these, "helpful" notifications in our lives have become a constant flow of interruptions, which are increasingly difficult to manage (Hayes & Miller, 2010). One potential solution is to investigate how the timing of an interruption influences disruptiveness.

Numerous studies have shown that interrupting an individual at random moments while completing a task can lead to a decrease in performance on the primary task (Bailey & Konstan, 2006) as well as an increased feeling of frustration, anxiety, and annoyance (Adamczyk & Bailey, 2004). In fact, when a system has poorly implemented interruptions, they are considered more of a nuisance than helpful (Dorneich, Ververs, Mathan, Whitlow, & Hayes, 2012), and can result in individuals refusing to use the system altogether (Hayes, Pande, & Miller, 2002). Conversely, when a system executes an interruption at an appropriate moment, the user can complete the task faster and with fewer errors (Bailey & Konstan, 2006), which can also foster trust in the system (Adamcyzk & Bailey, 2004; Iqbal & Bailey, 2005). Therefore, identifying the appropriate moment for an interruption is of utmost importance. One approach to identifying opportune moments for an interruption is to consider an individual's level of mental workload (Miyata & Norman, 1986).

Mental Workload

Mental workload has been defined as the interaction of a task's demands and an operator's processing capacity (Hart & Staveland, 1988). Specifically, task demand entails the coordination of mental and physical resources needed to complete a task (Jex, 1988). Although mental workload sometimes includes a component of physical workload, it differs from an outright measure of physical activity. Physical activity reflects the relationship between the amount of work completed and the cost of energy expended to perform that amount of work (Reid & Nygren, 1988). Therefore, physical workload is concerned with the amount of physical work required to complete a task, whereas physical activity is concerned with the amount of muscle fatigue required to complete a task (Jex, 1988).

Due to the complex and multifaceted nature of mental workload (Gopher & Donchin, 1986; Jex, 1988), several considerations have been proposed to assess the effectiveness of a workload measurement technique (Wierwille & Eggemeier, 1993). The two most important considerations include sensitivity and intrusiveness (Eggemeier, 1988). Sensitivity is the degree to which the measure can distinguish among different levels of workload. Intrusiveness represents the degree to which implementing the technique impacts the operator's performance. Ideally, a workload technique should have high sensitivity and low intrusiveness. Approaches to measuring mental workload typically fall into three broad categories: performance indices, physiological indices, and subjective estimates (Eggemeier, 1988; Meshkati & Loewenthal, 1988; Wierwille & Eggemeier, 1993).

Performance on the primary task of interest is often used as an indicator of underlying mental workload (Hockey, 1997) because it reflects the interaction between the mental demands of the task and capabilities of the operator to handle these demands (Cain, 2007). Mental workload associated with an easy task is generally low because excess resources are available. For a difficult task, however, mental workload is high because all resources are dedicated to the task and no additional resources are available should task demand increase further (Kahneman, 1973). Some typical performance measures used to reflect mental workload include accuracy, reaction time, and the number of errors (Cain, 2007; Paas & Van Merrienboer, 1993). The advantage of performance measures is that they are nonintrusive and are potentially available for any task. However, findings related to the sensitivity of this approach have produced mixed results (Wierwille & Eggemeier, 1993), because when task demands are too low using few resources or too high using all resources, performance may remain unchanged with shifts in workload (Eggemeier, 1988).

An alternative performance measure of mental workload uses a secondary task (Hart & Staveland, 1988; Verwey & Veltman, 1996), which assesses an operator's residual capacity while performing the primary task (Cain, 2007). According to the capacity theory of attention (Kahneman, 1973), performance is expected to decline when two tasks compete for limited attentional resources. For example, when performing two tasks simultaneously, there is an increase in working memory demands and the cost of switching between tasks, which increases mental workload and can decrease performance (Solovey et al., 2011). The advantage of using this approach is that it has high sensitivity if the secondary task resources overlap with those of the primary task (Wierwille & Eggemeier, 1993). However, a disadvantage to this approach is that introducing a secondary task can be highly intrusive.

Physiological indices are thought to reflect mental workload by assessing physiological arousal (Rubio, Diaz, Martin, & Puente, 2004). Several physiological indices used to measure mental workload include heart rate variations, functional near-infrared imaging (fNIR), galvanic skin response, respiration, electroencephalogram (EEG), event-related potentials (ERP), oculomotor activity (EOG), and pupillometry (Bailey & Iqbal, 2008; Berka et al., 2007). Physiological indices are often utilized because they offer the advantage of monitoring performance in a continuous and dynamic way (Berka et al., 2007; Iqbal, Adamczyk, Zheng, & Bailey, 2005). However, continuous measurement is also considered a disadvantage to many physiological methods because it can be difficult to parse fluctuations due to differences in workload from other factors (Jacob & Karn, 2002). For example, even small variations in environmental conditions, such as a change in ambient light or screen luminance, can influence pupil dilation (Bailey & Iqbal, 2008). Additionally, the sensitivity and intrusiveness of this

approach are dependent on which physiological indices are utilized (Wierwille & Eggemeier, 1993).

The third approach used to measure mental workload is to ask participants to provide an estimate of their perceived level of task difficulty (Berka et al., 2007). Several subjective workload measures have been developed over the years, but the NASA-TLX (Hart & Staveland, 1988), SWAT (Reid & Nygren, 1988), and Cooper-Harper scale (Cooper & Harper, 1969) are among the most widely cited (Rubio et al., 2004; Yakobi, 2018). Subjective rating scales can be either single or multi-dimensional, the latter of which requires judgments about several psychological variables (Hart & Staveland, 1988). For example, the NASA-TLX produces a composite score of mental workload based on subscales that address mental demand, physical demand, temporal demand, performance, effort, and frustration. Compared to other workload assessment measures, subjective estimates offer the advantage of ease of implementation and nonintrusiveness (Rubio, et al., 2004; Wiereille & Eggemeier, 1993). However, a disadvantage to this method is that reporting is dependent on a participants' memory of workload, which can rapidly decay with time (Hart & Staveland, 1988). Yet, subjective measures of workload tend to be highly correlated with one another, enhancing the reliability of the subjective approach (Rubio et al., 2004; Verwey & Veltman, 1996).

Mental workload has been investigated with respect to the disruptiveness of interruption timing from two perspectives. The workload aligned task model uses physiological measures of mental workload to predict optimal interruption timing, whereas event segmentation theory relies on subjective mental workload measures to evaluate the disruptiveness of interruptions based on timing. The following sections will discuss how each approach utilizes mental workload to determine the optimal moment for an interruption to occur.

Workload-Aligned Task Model

The workload aligned task model is a technique for determining optimal interruption timing using mental workload measures (Bailey, Busbey, & Iqbal, 2007; Bailey & Iqbal, 2005; Iqbal, et al., 2005; Iqbal & Bailey, 2005). According to this model, the best moment to interrupt a user corresponds to a period of low mental workload. Conversely, the worst moment for an interruption corresponds to a period of high mental workload. These predictions are based on an influential statement made by Miyata and Norman (1986), who argued that notifications delivered at moments of lower mental workload would have lower interruption cost; and that these moments occur between the completion of one subgoal and the start of another subgoal, known as a subtask boundary. This juncture is expected to result in lower workload because attentional resources from the previous subtask are released and there is a short lull before attentional resources must be refocused on the next subtask. Several studies provide evidence for these predictions.

Iqbal, et al. (2005) measured fluctuations in mental workload via pupillometry as participants completed a route planning and document editing task. Specifically, the researchers were interested in how workload changes with respect to subtasks and subtask boundaries. Subtask boundaries were defined as the time between the completion of one task and the start of another task. To determine the relationship between mental workload and subtask boundaries, the researchers decomposed each task into different levels using the goals, operators, methods, and selection rules (GOMS; Card, Moran, & Newell, 1983) task analysis technique. The researchers identified the overall goal of the task, then decomposed that goal into subgoals and operators until there was no observable separation between subgoals or operators remaining. The overall goal of the task represents the highest level in the hierarchy and each decomposition of a

goal into a subgoal is moving down one level in the hierarchy. Pupillary responses were then examined to understand how workload changed with respect to subtask boundaries. The results showed that across tasks, mental workload was found to be lower at the subtask boundaries. Further examination of the subtask boundaries revealed that workload decreased more for subtask boundaries higher in the task model than for boundaries lower in the task model. In other words, mental workload was the lowest for boundaries at the highest level in the hierarchy, and for each step down in the hierarchy, workload increased. The researchers additionally discovered that changes in mental workload among subtask boundaries within the same level of a task were inconsistent. This was one of the first studies to support the workload aligned task model.

Extending their previous work (Bailey & Iqbal, 2005; Iqbal et al., 2005), Bailey and Iqbal (2005) were interested in how executing interruptions at moments that produced the least and greatest amount of mental workload would impact interruption cost. The researchers selected subtask boundaries with the lowest and highest workload from route planning, document editing, and email classification tasks. While participants completed each of these tasks, they were interrupted at the best, worst, or random moments which were derived from the workload-aligned models developed in previous studies (Bailey & Iqbal, 2005; Iqbal et al., 2005). Dependent variables were subjective mental workload, resumption lag, annoyance, and how respectful users perceived the system (i.e., social attribution). The results indicated that across tasks, there was less disruption overall when a user was interrupted at the best boundary than any other boundary. This was characterized by a shorter resumption lag, decreased feelings of annoyance, and increased feelings of social attribution. However, unexpectedly the timing of the interruption showed no effect on subjective reports of mental workload. This is an interesting finding because it means that their physiological measures of mental workload did not

correspond to the subjective measures of mental workload. Yet, the results were some of the first to show that interruptions during periods of lower mental workload reduced interruption cost. To gain a better understanding of their results Bailey and Iqbal (2008) performed a different analysis of their data.

Using the same GOMS technique, Bailey and Iqbal (2008) decomposed the document editing, route planning, and email classification tasks used in Bailey and Iqbal (2005). Participants then completed each of these tasks while changes in pupillary response were recorded. After completion of the task, the researchers aligned pupillary responses to subtask and subtask boundaries, derived from their task analysis. This alignment process differed from the first analysis due to several improvements that were thought to more accurately capture the changes in mental workload. Consistent with the previous analysis (Iqbal et al., 2005), subtasks induced greater mental workload compared to baseline mental workload measures. The results also showed that mental workload was lower at subtask boundaries than within subtasks, although this difference was not significant. Contrary to previous findings, the period of lowest mental workload corresponded to a boundary in the middle of the task hierarchy, although this was previously found to occur at the highest boundary in the model (Iqbal et al., 2005). The inconsistency between studies has never been addressed and no study to date has tested these revised breakpoints to determine if they are more beneficial than previously identified breakpoints.

Although these studies (Bailey & Iqbal, 2005; Bailey & Iqbal, 2008; Iqbal, et al., 2005; Iqbal & Bailey, 2005) offer some support for using workload as a predictor of opportune moments for an interruption, there are several concerns with this approach. Perhaps the most significant concern is the lack of consistency across studies regarding the optimal breakpoints.

Without consistency, it can be difficult to understand what produces lower mental workload at these moments, why these moments result in less disruptive interruptions, and ultimately where future interruptions should occur to reduce disruptiveness. Additionally, these researchers never operationally defined low and high workload, which adds to the inconsistencies across studies. A final concern surrounding this method is that the physiological measures of mental workload did not correspond with subjective measures, which have been shown to be reliable (Hart & Staveland, 1988). This may indicate that physiological measures are not reliably capturing changes in mental workload, calling into question the results of workload-aligned task model studies. Based on these concerns, another technique used to study interruption timing, event segmentation theory (Zacks, Tversky, & Iyer, 2001), focuses on natural breakpoints in a task as an indicator of optimal interruption timing.

Event Segmentation Theory

Event segmentation theory (EST) developed by Zacks, Speer, Swallow, Braver, and Reynolds (2007) explains that events are perceived and encoded into discrete episodes. An event is a segment of time, perceived to have a beginning and an end, and occur in a given location (Zacks & Tversky, 2001). Zacks et al. (2001) explain that a task can be divided into coarse and fine segments based on its hierarchical representation in the mind. Coarse segments represent salient parts of a task such as the introduction of new information or objects (Horton & Rapp, 2003) or actions taken on an object, while fine segments represent small units within a task that are considered to be meaningful (Adamcyzk & Bailey, 2004; Zacks et al., 2001). For example, when participants provided descriptions of the course and fine segments associated with making a bed, putting on a sheet was classified as a course segment, while opening a drawer, taking the sheet out, and unfolding the sheet were classified as fine segments (Zacks, Tversky, & Iyer,

2001). According to EST, the best moment for an interruption to occur is between two coarse segments, known as a coarse breakpoint, because it is better recalled and understood (Adamczyk & Bailey, 2004).

In a study of EST, Adamczyk and Bailey (2004) interrupted a user at coarse, fine, and random times during a web searching, document editing, and media tasks to understand the impact of interruption timing on a performance. Results from this study indicated that the best moment for an interruption was during a coarse breakpoint, characterized by a reduction in frustration, annoyance, time pressure, resumption lag, mental demand, and mental effort.

Bailey and Konstan (2006) also examined the effects of interruptions at different boundaries. While completing a variety of tasks, participants were either interrupted between two tasks, during a task, or not at all. Interruptions occurring at coarse breakpoints were predicted to result in less disruption, measured by task performance, error rate, task time, task difficulty, annoyance, and anxiety. The results indicated that when an interruption occurred between tasks, participants reported lower annoyance and anxiety. In addition, task time was longer with interruptions than without interruptions; however, this prolonged time was eliminated when the interruption occurred at coarse breakpoints. Although the results of these studies support the idea that interruptions during coarse breakpoints are less disruptive, there is some debate about the accuracy of how breakpoints are determined.

The traditional approach to determining breakpoints comes from early studies done by Newtson (1973). In these studies, participants were instructed to watch a video of someone performing a task while they indicated, "in their judgment, where one unit ends and another begins" (p.32). Using this approach, breakpoints are determined on an individual basis, which has been shown to remain largely consistent across time and individuals (Zacks et al., 2007). In

addition, the division of a task into coarse and fine segments has been reliably detected by fMRI brain activity, where transient changes in activity in the posterior and frontal cortex correspond to previously identified event boundaries, and this activation is greater for course segments (Zacks & Sargent, 2010). Despite the consistency reported in these studies, there is equally strong evidence suggesting that this approach for determining breakpoints can be easily manipulated.

Zacks and Tversky (2001) found that participants' previous knowledge and goals can influence their individual level of segmentation. To examine this idea, Graziano, Moore, and Collins (1988) had graduate students view a video of a child completing a task and divide it into segments. Results of the study indicated that when graduate students were provided background information about the child in the video, such as pets they own or things they like, they segmented the video into significantly larger chunks. In addition, Zacks and Tversky (2001) found evidence suggesting that participants' degree of segmentation is influenced by the instructions they are given. Considering the variability in responses and the degree to which they can be influenced, it is surprising that most studies use only 60-80% individual agreement to determine breakpoints (Adamczyk & Bailey, 2004). Despite these concerns, this process is the predominant method for determining breakpoints using EST (Kurby & Zacks, 2008; Zachs, 2004; Zacks et al., 2001).

EST builds on the work of the workload-aligned task model by offering a more objective and consistent approach to determining breakpoints by utilizing task boundaries compared to physiological measures of workload. Yet both EST and the workload aligned task model support the idea that interruptions occurring at different boundaries are not equally disruptive, with

neither approach offering a theoretical foundation to explain why these differences are observed. Therefore, the present study turned to memory for goals theory as a possible explanation.

Memory for Goals Theory

Memory for goals (MFG) theory based in the ACT-R framework (Anderson, 1996) has been used to explain the cognitive process of how people remember, retrieve, and forget goals on which they have worked (Trafton et al., 2013). According to this theory, when an interruption occurs, the goal of the primary task begins to decay and becomes difficult to remember. Memory has been shown to decay very rapidly even with interruptions that last half of a second (Monk, Bohem-Davis, & Trafton, 2002). Although decay occurs rapidly, there are strategies that can reduce the amount of decay.

Memory for goals can be strengthened with rehearsal or association with mental or environmental cues (Trafton et al., 2013). Evidence of rehearsal strengthening memory for goals comes from a study conducted by Salvucci, Taatgen, and Borst (2009). In their study, participants were either presented with an immediate interruption or an interruption with an eight-second interruption lag, while time to resume the primary task (resumption time) was recorded. Participants who were given an eight-second interruption lag had a significantly shorter resumption lag compared to participants who received an immediate interruption. According to the researchers, participants who experienced the eight-second interruption lag used that time to rehearse the goal of the task they were working on prior to the interruption, resulting in a shorter resumption lag. Consequently, participants who experienced the immediate interruption had no time to rehearse and therefore failed to retrieve the goal requiring re-encoding, which ultimately lengthened the resumption lag.

The influence of re-encoding on interruption disruptiveness is not well understood compared to the influence of memory retrieval. The process of re-encoding cues from the primary task environment to reorient the participant to what they were doing immediately prior to the interruption is called reconstruction. Salvucci (2010) examined the process of reconstruction after interruptions occurring at different breakpoints within a task. These breakpoints were consistent with predictions from EST theory, where the least and most disruptive interruptions were expected to occur at higher-level (coarse) and lower-level (fine) boundaries, respectively, within the task. The results indicated that the best boundary resulted in a six-second shorter resumption lag compared to the worst boundary. Salvucci argues that this difference is too large to be explained by memory retrieval alone and that reconstruction additionally impacts resumption time. At higher-level boundaries (coarse breakpoints), no reconstruction is needed, while at lower-level boundaries (fine breakpoints), several steps are required to reconstruct the task environment, contributing to a longer resumption lag.

An additional factor that impacts retrieval is the volume of information that needs to be retrieved from memory (Salvucci, Taatgen, & Borst, 2009). Retrieving more information from memory takes more time than less information, resulting in a longer resumption lag. Additionally, as the amount of information increases, so does the likelihood of failing to retrieve this information, leading to errors. The amount of information to be remembered is dependent on the level of representation of the primary task. According to the ACT-R framework, a task is represented as a hierarchy, comprised of higher to lower levels (Salvucci, 2010; Salvucci, Taatgen, & Borst, 2009). Like task decomposition described in workload aligned task models, the highest level in the hierarchy is the overall goal of the task, and each time the goal is divided into subgoals it creates lower levels in the hierarchy (Iqbal et al., 2005). Therefore, each lower

level is an accumulation of all levels above it, which provides a fuller representation of the task. However, this fuller representation requires more memory to be remembered. Accordingly, the researchers argue the best time to interrupt a user is at a higher-level subtask and the worst time to interrupt a user is at a lower-level-subtask. These predictions are in line with EST; however, MFG theory goes a step further by offering an explanation for why an interruption at a particular moment would influence disruptiveness. Specifically, MFG theory explains that at higher levels, the user's primary task representation is relatively low, meaning there is less to be remembered, which is predicted to shorten resumption lag and produce fewer errors (Salvucci, Taatgen, & Borst, 2009). Subsequent studies have found empirical support for these predictions.

Trafton, Jacobs, and Harrison (2012) built and evaluated a model for interruption resumption based on MFG theory using the ACT-R framework. In the first experiment, participants were instructed to tell a story while being interrupted unexpectedly. The researchers were interested in where participants resumed after the interruption and the length of the resumption lag. The results indicated that participants who were interrupted took longer to resume the primary task compared to participants who were not interrupted. Additionally, participants often resumed by repeating what they had already said, rather than beginning at the correct location or skipping ahead. When this result was compared to predictions made by the model, it showed an excellent fit to the experimental data. To further validate their predictions, the researchers ran another storytelling experiment in which the model was used to determine how accurately it could predict when a participant would need help remembering where they had left off. It was hypothesized that if the model can help in the resumption process, it is an accurate description of the cognitive processes underlying interruption resumption. The researchers found that the model was highly successful at predicting interruption resumption. Although this is a

good step forward in understanding the impact of interruption disruptiveness, it is important to evaluate this model when considering other features of interruptions, such as their frequency.

Interruption Frequency

Interruptions rarely occur in isolation, for example in the emergency room setting, interruptions occur as frequently as every three minutes (Healy et al., 2007). Interruption frequency; that is, the number of interruptions that occur when trying to complete a primary task (Jett & George, 2003; Lee & Duffy, 2015), or the time interval between interruptions (Galluch, Grover, & Thatcher, 2015; Monk, 2004; Tams, Thatcher, Grover, & Pak, 2015; Xia & Sudharshan, 2002).

Monk (2004) conducted one of the only studies examining the effect of frequency on interruption disruptiveness. Participants completed a pursuit-tracking task and were interrupted at either a low (every 30 seconds) or high (every 10 seconds) frequency. Monk postulated that resumption performance would be worse when there were more frequent interruptions. This prediction was based in MFG theory, which states that each time an interruption occurs a goal must be suspended, maintained, and resumed. Consequently, more interruptions should lead to greater goal retrieval interference, which includes longer resumption time, longer task completion time, and a greater number of resumption errors. Unexpectedly, Monk found that a high frequency of interruptions resulted in shorter resumption and task completion times with fewer errors compared to the infrequent condition. Although counterintuitive, several studies have similarly found a decrease in task completion time when interruption frequency is high (Speier, Valacich, & Vessey, 1999; Xia & Sudharshan, 2002; Zijlstra, Roe, Leonora, & Krediet, 1999).

A possible explanation is that when interruptions become too frequent, participants become so frustrated that they give up on the task and finish it as quickly as possible. Support for this idea comes from a study conducted by Lee and Duffy (2005), where they discovered in a pilot study that any more than three interruptions during a cognitive task resulted in a severe decrease in task performance. This decrease was due to frustration, annoyance, and lack of motivation, rather than the effects of the interruption itself. Supporting this idea further, the researchers conducted a larger-scale study and limited the high interruption frequency condition to three interruptions based on the results of the pilot. The researchers then measured how low, moderate, and high frequency interruptions impacted interruption disruptiveness. Despite limiting the high frequency condition to three interruptions, there was still a large effect on disruptiveness. Specifically, compared to the low (one interruption) and moderate (two interruptions) frequency conditions, the high frequency condition resulted in the highest error rates and longest completion time. This result is in line with the predictions of MFG theory, where the high frequency condition resulted in an increase in task completion time and the number of errors.

An alternative explanation for the decrease in completion time is that participants compensate for the increase in interruption frequency by spending less time on the primary task (Zijlstra et al., 1999). However, a consequence of compensating is that participants report higher mental workload to complete the task. This finding is consistent with the results of several other studies (Basoglu, Fuller, & Sweeney, 2009; Tams et al., 2015). In fact, Wickens and McCarley (2008) report that as the number of interruptions increases, mental resources become increasingly exhausted. Some researchers have even speculated that the increase in mental workload further impacts performance, resulting in more errors (Zijlstra et al., 1999).

A final possible explanation for these unexpected findings is that the timing of interruptions was random, rather than corresponding to optimal moments as predicted by MFG theory. However, no study to date has investigated the combined effects of interruption timing and frequency on disruptiveness, making the present study the first to do so.

The effect of frequency on interruption disruptiveness is not well studied, even less so in the MFG context (Morgan et al., 2009); however, some general trends do emerge. Although high, low, and moderate interruption frequencies are inconsistently defined across studies, as the frequency of interruptions increases, so does the cost to performance and the participant. Specifically, high interruption frequency is associated with an increase in the number of errors, mental workload, frustration, and task completion time. These general findings served as the foundation for understanding the impact of interruption frequency in the current study.

Goals of the Present Study

The present study applied EST to understand how interruptions executed at various breakpoints within a task affect disruptiveness. EST offers a more robust method for determining breakpoints because this method is not dependent on changes in mental workload, which often differ from one individual to another (Iqbal et al., 2005). Rather, EST relies on breaking down a task into coarse and fine segments, which may be more consistent across all users. However, even this approach to determining breakpoints has shown to be influenced by individual differences in segmentation (Graziano et al., 1988; Zacks & Tversky, 2001). Therefore, the present study applied hierarchical task analysis (HTA) to obtain a more objective representation of breakpoints. Using HTA, the primary task was decomposed into multiple subtasks until there were no remaining subdivisions, and then the subtasks were aligned to boundaries according to EST. To the author's knowledge, this is the first time HTA has been used within EST.

Additionally, there have been no studies to date that have tested the combined effects of interruption frequency and timing on disruptiveness. In order to test the complex nature of interruptions, the present study also investigated how well EST predicts the best moments for interruptions when they occur with low and high frequencies.

Finally, the present study sought to understand why the timing of an interruption matters. Currently, research on interruption timing focuses on measuring the cost associated with interrupting a user at specific moments but provides no insight into why the cost is higher/lower at those particular moments. MFG theory based in the ACT-R framework provides a possible explanation. Using information about the underlying cognitive processes that influence planning behavior, we can garner insight into the cause of these changes in perceived and actual interruption cost.

A coarse breakpoint was defined as the segmentation of a task, typically triggered by the start of a new action. These breakpoints were determined by decomposing tasks into hierarchical representations and selecting the highest level of segmentation for introducing an interruption. A fine breakpoint was defined as smaller actions within a task that are meaningful, which were also determined by selecting the lowest level of segmentation. An interruption was defined by the core attributes described by Brixey et al. (2007): intrusion of secondary, unplanned, and unexpected task; discontinuity; externally initiated; and situated within a context. In the present study, a pop-up message on a computer screen served as the interruption while an individual completed a task. The unexpected pop-up messages were secondary to a primary trip planning task. Additionally, the pop-up messages covered the entire screen, forcing the user to discontinue the primary task and immediately attend to the interruption. Finally, the pop-up messages were initiated externally by the researcher in a human-computer interaction context.

As suggested by EST and MFG, interruptions at fine breakpoints are more disruptive than interruptions at coarse breakpoints. When interrupted at a fine breakpoint, MFG theory predicts that the task representation in the interuptee's mind will be broader and therefore place greater demands on memory. Consequently, as memory demand increases so does the likelihood of forgetting (i.e., an error). Therefore, it was predicted that interruptions which occur at coarse and fine breakpoints would differentially impact the number of errors made. Specifically, it was predicted that more errors will be made when an interruption occurs at a fine compared to a coarse breakpoint.

Additionally, MFG predicts that interruptions at fine breakpoints will result in longer resumption times than at coarse breakpoints because the memory demands are proportional to the length of time it takes to resume. Therefore, it was predicted that the timing of an interruption will have different effects on the duration of the resumption lag due to differences in task representations held in memory. Specifically, it was predicted that resumption lag will be longest when an interruption occurs at a fine breakpoint compared a coarse breakpoint.

Consequently, after resuming the task, unlike coarse breakpoints where there is no need to re-encode the task environment, several steps of re-encoding are necessary after an interruption at fine breakpoints. As a result, it was hypothesized that task completion time will be longer to account for this re-encoding. Specifically, it was predicted that task completion time will be longest when an interruption occurs at a fine breakpoint compared to a coarse breakpoint.

Furthermore, it was hypothesized that mental workload will increase as more resources are needed to sustain the larger task representation held in memory at fine breakpoints. Therefore, it was predicted that interruption timing will differentially impact mental workload due to the differences in resources needed to maintain the task representation during an

interruption. Specifically, it was predicted that interruptions at fine breakpoints will result in higher subjective reports of mental workload compared to interruptions at coarse breakpoints. Also, the need to address interruptions at fine breakpoints should be perceived as more frustrating. Therefore, it was predicted that interruptions at fine breakpoints will result in higher levels of frustration compared to interruptions at coarse breakpoints.

In addition, the present study also sought to understand how well EST predicts opportune moments for an interruption when there are few or many interruptions. Research has shown when an interruption occurs, it results in longer task completion times. Consequently, because each interruption is associated with an incremental increase in task completion time, many interruptions will have an additive effect on task completion time. Therefore, it was predicted that the frequency of interruptions will differentially impact task completion times. Specifically, it was predicted that participants in the high frequency condition will experience the longest task completion times compared to participants in the low frequency condition.

Additionally, according to MFG a goal must be suspended, maintained, and resumed after an interruption. This process must be repeated for each interruption that occurs, meaning as the frequency of interruptions increases, so does the number of times an individual must undergo this process. Consequently, it was hypothesized that the more times an individual completes this cycle of suspending, maintaining, and resuming the task, the greater chance there is for making an error. Therefore, it was predicted that the frequency of interruptions will produce different numbers of errors. Specifically, it was predicted that more errors will be made when interruptions occur at a higher compared to lower frequency.

Furthermore, frequent interruptions have been found to increase mental workload and frustration levels. Therefore, it was predicted that a higher as compared to a lower frequency of interruptions will lead to higher levels of subjective mental workload and frustration.

Finally, although no study to date has examined the relationship between interruption timing and frequency, the present study was the first to conduct an exploratory analysis to understand if the effect of frequency on disruptiveness was influenced by whether an interruption occurs at a coarse or fine breakpoint.

Given that a higher frequency of interruptions is predicted to result in longer task completion times compared to a lower frequency, and that interruptions at fine breakpoints are predicted to result in longer task completion times than at coarse breakpoints; it was predicted that interruption timing and frequency will interact. Specifically, the combination of more frequent interruptions occurring at fine breakpoints will result in greater increases in task completion time, more errors, and higher levels of subjective mental workload and frustration.

CHAPTER II

METHOD

Participants

A power analysis was conducted using G*Power (Version 3.1.9.2) with a power of .80 and a medium effect ($f = 0.25$) with $\alpha = .05$ to determine an appropriate sample size for the study. The results indicated that a total of 66 participants should be recruited. All participants were at least 18 years of age ($M = 20.3$), had no history of living in Chicago in the past 5 years, and had normal or corrected-to-normal vision. Most participants were female (90.7%), and several participants were male (9.3%). All participants provided written consent prior to participating and were compensated with 1.5 psychology credits.

Design

The present study employed a 2 X 3 split-plot design. The within-subjects factor was interruption timing, which had three levels: none, coarse breakpoint, and fine breakpoint. The between-subjects factor was interruption frequency, which had two levels: few (1 interruption) and many (3 interruptions). The order of the task trials was counterbalanced across participants. The primary dependent variables of interest included resumption lag, task completion time, errors, frustration, and mental workload.

Primary Task

The primary task required participants to complete an unfinished itinerary for a weekend trip (see appendix E). Participants began by reading a passage explaining that their friend Anne was planning a weekend trip to Chicago when she suddenly got swamped at work and no longer has time to finish planning. Participants were tasked with using the flight and additional

information Anne left behind to finish planning the trip. To accomplish this, participants completed three task trials, one for each of the following tasks: selecting a rental car, creating an itinerary, and selecting a restaurant, while maintaining an estimate of trip expenses.

Selecting a rental car. The first task trial was to select one of four predetermined vehicle types. To do this, participants had to research each vehicle type on the Enterprise rental page using the link provided. From the website, participants gathered information about the size, look, and price of each vehicle which was then used to decide which car was better suited for the needs of the trip.

Creating an itinerary. The second task trial was to gather information about three predetermined activities and use that information in combination with flight information to create an itinerary of when each activity should be scheduled. Using the links provided, participants researched hours of operation, price, and any additional information they deemed necessary to plan an itinerary for the Brookfield Zoo, Museum of Science and Industry, and the 360 Chicago Observation Deck.

Selecting a restaurant. The final task trial was to research four predetermined restaurants and then select the one that met the needs of the group. This included selecting a restaurant that was open on the night of arrival in Chicago, with a reasonable price, and several vegan options. Using the links provided, participants navigated each restaurant website for the menu, price, and hours of operation. Participants then used that information to inform their decision.

Hierarchical Task Analysis

To establish a more objective approach for determining breakpoints, HTA was used. Figure 1 shows a broad overview of the HTA for the select a rental car task. (The detailed HTA is written in a list format rather than a visual format and can be found in Appendix D.) The overall goal of the task was to select a rental car, which is represented as the top box in the hierarchy. The overall goal was divided into the steps a user must perform to accomplish this goal, known as subtasks. The period between adjacent subtasks is referred to as the subtask boundary. The level of each subtask and subtask boundary is tied to the number of divisions from the primary goal. For example, to complete the overall goal of selecting a rental car, the participant researched each of the four vehicles and then selected one. These subtasks represent the first level in the task model because they are the first division of the goal into subtasks needed to accomplish that goal. Consequently, boundaries at this level are known as level-one boundaries.

From this initial division, each of these subtasks was further divided into additional subtasks until there were no remaining subdivisions. Each decomposition moves a level lower on the hierarchy. For example, to research a vehicle, the participant started a reservation, chose a vehicle class, selected extras, and reviewed the vehicle price. These subtasks are two levels below the overall goal and therefore represent the second level in the task model. Consequently, boundaries at this level are known as level-two boundaries. To ensure reliability in the resulting hierarchies, each task trial was decomposed in this manner on two separate occasions. Inconsistencies between the two hierarchies were remedied by decomposing that task trial on a third occasion and then selecting the hierarchy that was consistent on at least two of the occasions.
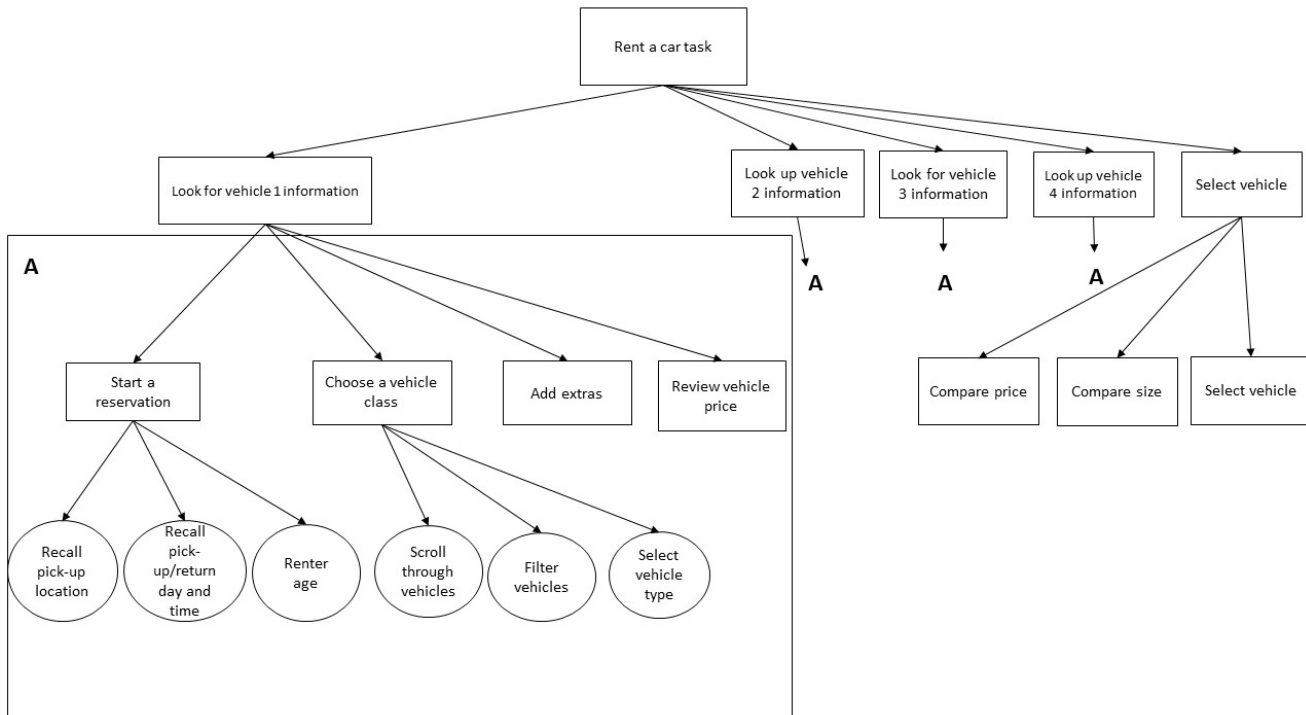
Figure 1. A broad overview of the HTA for the select a rental car task trial. The bold **A** represents a part of the task that is repeated elsewhere in the task analysis.

Overviews of the task analyses for the other two task trials (creating an itinerary and selecting a restaurant) can be found in Appendix C. In addition to these hierarchies, plans were created to define the order in which the subtasks may be performed to complete the overall task. The plan for each task trial can be found in red ink in Appendix D. Decomposing a task in this manner has been applied to the workload-aligned task model, but to the author's knowledge, this method has never been applied to EST.

After the HTA was completed, subtasks were aligned to coarse or fine boundaries in accordance with EST. The top of the hierarchy is the most coarse task while the bottom row of the hierarchy represents the finest segments within a task. Thus, the hierarchy goes from coarse to fine segments, but with no formal distinction as to where segments change from coarse to fine. To account for this ambiguity, the most coarse and fine segments within the hierarchy were used

to establish the coarse and fine breakpoints for the present study. However, because there is a different number of levels for each task, the lowest level was always used across all tasks for the fine breakpoints. These breakpoints are shown in Table 1. A pilot study was conducted that further validated the hierarchical task analysis and selected breakpoints.

Table 1

Description of interruption triggers by breakpoint and task type

Task Trial	Breakpoints	Interruption Trigger
Select a rental car	Coarse	1. After recording price of vehicle 1 and before clicking back to Enterprise to look for vehicle 2 information. 2. After recording price of vehicle 2 and before clicking back to Enterprise to look for vehicle 3 information. 3. After recording price of vehicle 3 and before clicking back to Enterprise to look for vehicle 4 information.
	Fine	1. After reviewing vehicle 1 extras before recording vehicle 1 price. 2. After reviewing vehicle 2 extras before recording vehicle 2 price. 3. After reviewing vehicle 3 extras before recording vehicle 3 price.

Table 1 Continued

Create an itinerary	Coarse	1. After recording activity 1 information and before clicking on link to research activity 2 information. 2. After recording activity 2 information and before clicking on link to research activity 3 information. 3. After recording activity 3 information and before clicking on new page to start itinerary order.
	Fine	1. After finding admission price for activity 1 before recording admission total. 2. After finding admission price for activity 2 before recording admission total. 3. After finding admission price for activity 3 before recording admission total.
Select a restaurant	Coarse	1. After recording restaurant 1 information before clicking on link to research restaurant 2. 2. After recording restaurant 2 information before clicking on link to research restaurant 3. 3. After recording restaurant 3 information before clicking on link to research restaurant 4.
	Fine	1. After looking at meal prices at restaurant 1 before recording average meal price. 2. After looking at meal prices at restaurant 2 before recording average meal price. 3. After looking at meal prices at restaurant 3 before recording average meal price.

Note. Interruption triggers represent when the interruption will be executed by the researcher based on observable actions by the participant. Interruption triggers are classified by breakpoints (coarse or fine) and by task trial (select a rental car, create an itinerary, or select a restaurant).

Task Interruptions

Throughout completion of the primary task, participants were unexpectedly interrupted. However, to ensure the participants were reading the interruptions, they were informed that several offers relevant to the itinerary may pop-up while they completed the task, and that these pop-ups should be read before closing the message. To ensure interruptions were executed at the appropriate moment, the researcher created a checklist for each participant that specified the key combinations needed to execute each interruption, as well as the participant action that signified when the researcher should introduce each interruption by task trial and condition (see appendix K for an example checklists). The interruptions were presented in the form of computer system alerts in pop-up messages displaying a suggestion for the primary task. The interruptions were executed via a secondary keyboard controlled by the experimenter. Using a program called AutoHotKey, interruptions were coded to be triggered and pop up in accordance with specific keyboard combinations. For example, hitting the "control" key and the "Q" key at the same time would initiate a pre-programmed interruption. A full list of interruptions per task trial can be found in Appendix H. All interrupting messages used were approximately 16 words in length.

The interrupting message was relevant to the task; however, it did not require additional action to be taken. For example, during completion of the itinerary portion of the task participants were presented with a message such as, "Admission prices are 20 percent off when you visit Monday-Thursday". This interruption is related to the task, however, because the trip takes place on a weekend, it did not require participants to alter the itinerary to accommodate the message. The message was presented on their displays and occluded the primary task, forcing the participant to switch their attention to the interruption. The timing of the interruption was

recorded by AutoHotKey as the time from when the interruption was presented on the screen, until the time the participant hit the X button at the top of the screen to close the message.

During the three task trials participants received (a) interruption(s) at either a course breakpoint, a fine breakpoint, or not at all. The number of interruptions that each participant received was related to whether they were in the few or many interruption condition. Participants in the many interruption condition received three interruptions for that task trial, which are displayed in Table 1. For example, a participant in the many interruption condition, completing the restaurant task received all three of the interruptions displayed in Table 1 to the right of the "select a restaurant" task. However, for some of the participants, the three interruptions were executed at moments that corresponded to coarse breakpoints, while for other participants, the three interruptions were executed at moments that corresponded to fine breakpoints, depending on which condition they were in. Conversely, participants in the few interruption condition received only one interruption. To determine which of the three interruptions participants would receive, interruption number (one, two, or three) was randomized across participants in the few interruption condition. For example, a participant in the few interruption condition, completing the restaurant task received only one of the three possible interruptions displayed in Table 1 to the right of the "select a restaurant" task. So, some of the participants received only interruption one, while others received only interruption two, and others received only interruption three. Similar to the many condition, the interruption number participants received was related to whether the task trial they were assigned was coarse or fine.

Measures

Resumption lag. Resumption lag refers to the time it took participants to switch from attending to the interruption to resuming the primary activity. This was measured as the time

from when the participant acknowledged they read the pop-up message, automatically recorded by AutoHotKey, to when the participant resumed activity on the primary task, recorded to the nearest tenth of a second. Specifically, resuming the activity was represented by the first relevant action taken toward the primary task. An example of a relevant action was clicking on the on-screen calculator for the purpose of calculating a total. An example of an irrelevant action was clicking on the word document to reference flight information needed to complete the task. The time of the first relevant action was derived from video recording of that task trial. Resumption lag has been measured in this manner across a variety of contexts (Adamczyk & Bailey, 2004; Altmann & Trafton, 2004; Bailey & Iqbal, 2005; Monk, 2004; Monk, Boehm-Davis, & Trafton, 2004; Trafton & Monk, 2007). Additionally, the researcher slowed the video speed to ¼ of the pace to ensure the most accurate resumption lag was recorded. Furthermore, the researcher recorded the resumption lag on two separate occasions to ensure reliability. Any inconsistent resumption lags between the two occasions was accounted for by re-coding that trial on an additional occasion and using the resumption lag that was consistent on at least two of the occasions.

Task completion time. Total time on the task was recorded along with the time it took for the participant to attend to the interruption. Task completion time was then derived from these numbers by subtracting the time it took participants to attend to the interruption from total task time, rounded to the nearest tenth of a second. Task completion time has been measured in this manner by previous studies on interruption timing (Adamczyk & Bailey, 2004).

The time taken to attend to the interruption includes the time of the interruption (measured by AutoHotKey and explained above), as well as the resumption lag (also explained above). The total time of the task was derived from video recording of that task trial.

Specifically, the start time was objectively recorded as the moment participants stopped scrolling from the intentionally blank page to the instruction page at the start of the task, and end time was objectively recorded as either the moment participants scrolled down from the end of the task worksheet or when participants verbally indicated they were done with the task, whichever occurred first. Similar to resumption lag, the video speed was slowed to ¼ of the pace to ensure accuracy and coded twice to ensure reliability. Inconsistencies were handled with the same process as detailed above for resumption lags.

Number of errors. The total number of errors was recorded for each participant. Errors included exceeding the predetermined budget, violating a restriction stated in the instructions (i.e., selecting a restaurant that is not open the night of arrival), and misusing or lacking information to complete a task (i.e., booking an activity after the departure flight time). The number of errors possible differed depending on each task trial.

There were two possible errors participants could make during the rental car task trial: 1) they could enter incorrect reservation information, which altered the cost of all the rental vehicles, or 2) they could record the incorrect cost of the selected vehicle. If participants made the first error, they were not further penalized for an incorrect total of the selected vehicle, as long as the cost of the selected vehicle matched the cost recorded for that vehicle on the previous page. There were eight possible errors participants could make during the itinerary task trial. Errors included incorrectly recording the admission price or hours of operation for the zoo, itinerary, and 360 observation deck, as well as an incorrect itinerary order or incorrectly recording the group activities total. Finally, there were nine possible errors participants could make during the restaurant task trial. Errors included incorrectly recording whether the restaurant

was open the night of arrival or if there were several vegan options on the menu for each restaurant, as well as incorrectly choosing the final restaurant.

Mental workload and frustration. Subjective workload was assessed by the NASA-TLX, a measure of perceived mental workload (Hart & Staveland, 1988). The NASA-TLX allowed participants to report measures of mental demand, physical demand, temporal demand, performance, effort, and frustration. Responses were recorded by circling a vertical line on a continuous scale ranging from low (1) to high (20). This scale has been used in previous studies to measure mental workload after interruptions (Adamcyzk & Bailey, 2004). The NASA-TLX has achieved high levels of test-retest reliability as well as a Cronbach's alpha of .80 (Hart & Staveland, 1988; Xiao, Wang, Wang, & Lan, 2005) and concurrent validity additional measures of workload (Rubio, Diaz, Martin, & Puente, 2004).

Procedure

Participants read and signed an informed consent form (see appendix A) and completed a background questionnaire (see appendix B). Participants were then randomly assigned to either the few or many interruption conditions. Within each of these groups, the order of selecting a rental car, creating an itinerary, and selecting a restaurant task trials was counterbalanced across participants. Prior to beginning the task, the experimenter read through an introduction script, which included the overall task instructions (see appendix J), to ensure participants understood the task. Participants then completed the primary task of planning a trip to Chicago while being interrupted in accordance with the condition they were in. After completing the task, participants completed the NASA-TLX survey (see appendix F). Participants repeated the process with the other two tasks. Once the responses were recorded on the last NASA-TLX participants were debriefed and thanked for their time.

CHAPTER III

RESULTS

A total of 59 participants were recruited from undergraduate psychology classes at Old Dominion University. Five participants were excluded from data analysis because the of technical glitches (i.e., camera stopped recording), for failing to follow task instructions (i.e., skipping a portion of a task), asking too many questions (more than 3 per task or 6 overall), or the researcher failed to execute an interruption at the appropriate moment. Thus, the final number of participants was 54. Although a power analysis suggested a sample size of 66 participants, the results of counterbalancing revealed that there were 27 unique condition combinations. From this observation, the researcher determined that a minimum of 54 participants would be recruited so that each unique condition was tested twice. If there was marginal significance with 54 participants, the researcher planned to recruit 27 more participants. However, the analyses suggested another 27 participants would not affect the interpretation of the results.

Prior to statistical analysis, all data were checked for outliers and a normal distribution and these criteria were met. Levene's test of homogeneity was used to determine that variability in performance was equal across all groups. Mauchly's test of sphericity was used to determine that the variability in the difference between all possible pairs of within-subject conditions were equal. Normality was assessed with a Shapiro-Wilks test. Any violations of normality were addresses via a natural logarithm transformation. Outliers were assessed via boxplots. Any outliers more than 1.5 standard deviations from the mean were replaced with the group mean. Comparisons among means were assessed with a Bonferroni correction.

Number of Errors

Results from a descriptive statistical analysis indicate that on average, individuals who received no interruptions made the fewest errors, followed by individuals who received a(n) interruption(s) at fine breakpoints, and then individuals who received a(n) interruption(s) at coarse breakpoints (see Table 2). A 2 (few, many) X 3 (coarse, fine, none) split-plot ANOVA was conducted to assess if there were differences in the number of errors an individual made on the itinerary task based on interruption frequency and interruption timing (see Table 3). The analysis revealed no significant effects.

Table 2

Mean Number of Errors Committed within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	n	M	SD
Coarse Breakpoint	Few	27	1.11	1.09
	Many	27	1.07	1.14
	Total	54	1.09	1.10
Fine Breakpoint	Few	27	0.96	0.94
	Many	27	1.00	0.78
	Total	54	0.98	0.86
No Interruption	Total	54	0.87	0.78

Table 3

Results of the Analysis of Variance for Number of Errors

	SS	df	MS	*F*	*p*	partial η^2
Interruption Timing	1.33	2	0.67	0.79	.456	.015
Interruption Timing x Interruption Frequency	0.35	2	0.17	0.21	.815	.004
Error	87.65	104	0.84			
Interruption Frequency	0.15	1	0.15	0.17	.683	.003
Error	47.46	52	.91			

Resumption Lag

Prior to statistical analyses three outliers were replaced with the group mean. Results from a descriptive statistical analysis indicate that on average, individuals in the fine breakpoint condition had longer resumption lags than those in the coarse breakpoint condition (see Table 4). To assess the difference in the length of resumption lag on the itinerary task due to interruption frequency and interruption timing, a 2 (few, many) X 2 (coarse, fine) split-plot ANOVA was performed (see Table 5). The analysis revealed a main effect of interruption timing on resumption lag. Specifically, resumption lag was significantly longer when participants were interrupted at a fine breakpoint compared to a coarse breakpoint.

Table 4

Mean Length of Resumption Lag within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	*n*	*M*	*SD*
Coarse Breakpoint	Few	27	3.02	1.49
	Many	27	3.28	1.18
	Total	54	3.15	1.34
Fine Breakpoint	Few	27	9.61	7.25
	Many	27	7.59	4.82
	Total	54	8.60	6.18

Table 5

Results of the Analysis of Variance for Resumption Lag

	SS	df	MS	*F*	*p*	partial η^2
Interruption Timing	802.08	1	802.08	38.56	**.001	.426
Interruption Timing x Interruption Frequency	35.27	1	35.27	1.70	.199	.032
Error	1081.75	52	20.83			
Interruption Frequency	20.87	1	20.87	1.11	.298	.021
Error	980.92	52	18.86			

Note. $* p < .05, ** p < .001$

Task Completion Time

Prior to running the analyses, a violation of normality was corrected with a natural logarithm transformation. Results from a descriptive statistical analysis indicate that on average, individuals who were interrupted at fine breakpoints had the longest task completion time,

followed by individuals who were not interrupted, and then individuals who were interrupted at coarse breakpoints (see Table 6). To assess the difference in the length of task completion time on the itinerary task due to interruption frequency and interruption timing, a 2 (few, many) X 3 (coarse, fine, none) split-plot ANOVA was performed (see Table 7). The analysis, however, revealed no significant effects.

Table 6

Mean Length of Task Completion Time within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	*n*	*M*	*SD*
Coarse Breakpoint	Few	27	667.52	209.55
	Many	27	654.56	189.48
	Total	54	661.04	197.98
Fine Breakpoint	Few	27	699.68	249.04
	Many	27	663.00	198.00
	Total	54	681.34	223.61
No Interruption	Total	54	674.59	179.56

Table 7

Results of the Analysis of Variance for Task Completion Time

	SS	df	MS	*F*	*p*	partialη^2
Interruption Timing	11540.33	2	5770.17	0.18	.832	.004
Interruption Timing x Interruption Frequency	68806.98	2	34403.49	1.10	.338	.021
Error	3265691.78	104	31400.88			
Interruption Frequency	499.98	1	499.98	0.01	.927	.001
Error	310128939	52	59640.18			

Mental Workload

To achieve a measure of mental workload, the subscales on the NASA-TLX were summed and the resulting value represents the total mental workload score for each participant. Results from a descriptive statistical analysis indicate that on average, individuals who received a(n) interruption(s) at fine breakpoints reported the highest total mental workload, followed by individuals who received a(n) interruption(s) at coarse breakpoints, and then individuals who received no interruptions (see Table 8). To assess the difference in total mental workload on the itinerary task due to interruption frequency and interruption timing, a 2 (few, many) X 3 (coarse, fine, none) split-plot ANOVA was performed (see Table 9). The analysis revealed a significant main effect of interruption timing on total mental workload.

Table 8

Mean Total Mental Workload Score within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	*n*	*M*	*SD*
Coarse Breakpoint	Few	27	50.93	17.24
	Many	27	51.98	15.51
	Total	54	51.45	16.25
Fine Breakpoint	Few	27	56.78	18.15
	Many	27	54.61	21.72
	Total	54	55.69	19.85
No Interruption	Total	54	47.91	16.82

Table 9

Results of the Analysis of Variance for Total Mental Workload

	SS	df	MS	*F*	*p*	partial η^2
Interruption Timing	1641.57	2	820.78	11.32	**.001	.262
Interruption Timing x Interruption Frequency	259.91	2	129.96	1.79	.172	.078
Error	7543.36	104	72.53			
Interruption Frequency	176.30	1	176.30	0.22	.642	.004
Error	41911.31	52	805.99			

Note. $* p < .05, ** p < .001$

Mean total mental workload scores are shown in Figure 2. Pairwise comparisons were conducted and indicate that total mental workload was significantly lower in the no and coarse interruption conditions when compared to the fine breakpoint condition. However, total mental

workload did not significantly differ between the no interruption condition when compared to the coarse breakpoint condition.
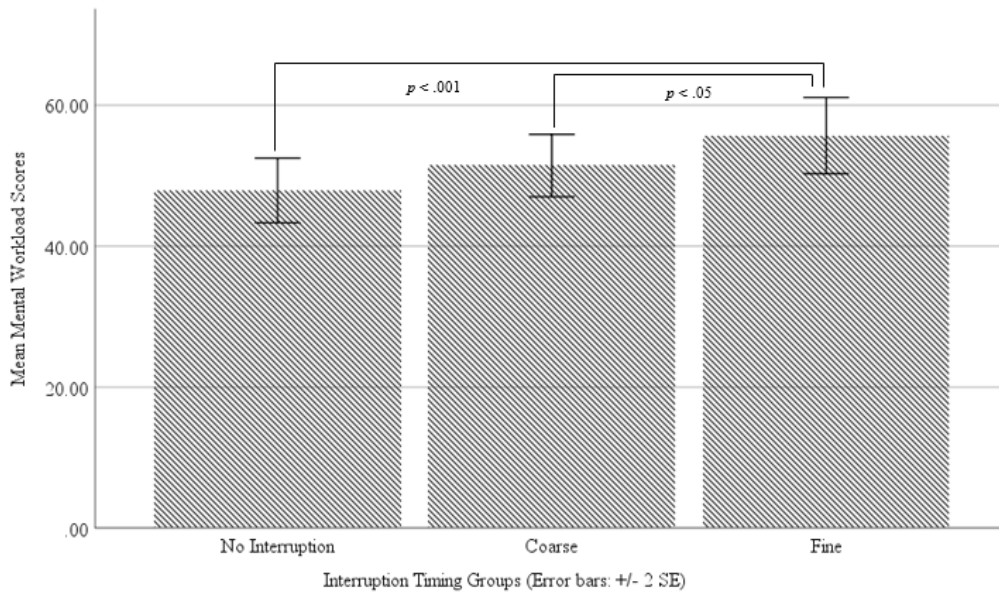
Figure 2. Mean total mental workload scores within interruption timing conditions. Error bars represent standard errors.

Frustration

To test if there were differences in frustration due to interruption timing or frequency, as predicted, the frustration subscale of the NASA-TLX was analyzed independently. Mean frustration scores are shown in Figure 3. Prior to running the analyses, a violation of normality was corrected with a logarithmic transformation. Results from a descriptive statistical analysis indicate that on average, individuals who received a(n) interruption(s) at fine breakpoints reported the highest frustration, followed by individuals who received a(n) interruption(s) at coarse breakpoints, and then individuals who received no interruptions (see Table 10). To assess the difference in frustration on the itinerary task due to interruption frequency and interruption

timing, a 2 (few, many) X 3 (coarse, fine, none) split-plot ANOVA was performed (see Table 11). The analysis revealed a significant main effect of interruption timing on frustration.

Table 10

Mean Frustration Score within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	n	M	SD
Coarse Breakpoint	Few	27	0.55	0.40
	Many	27	0.67	0.31
	Total	54	0.61	0.36
Fine Breakpoint	Few	27	0.71	0.31
	Many	27	0.70	0.42
	Total	54	0.71	0.37
No Interruption	Total	54	0.54	0.35

Table 11

Results of the Analysis of Variance for Frustration

	SS	df	MS	F	p	partialη^2
Interruption Timing	0.77	2	0.38	5.89	*.004	.102
Interruption Timing x Interruption Frequency	0.12	2	0.06	0.88	.417	.017
Error	6.78	104	0.07			
Interruption Frequency	0.08	1	0.08	0.31	.578	.006
Error	13.68	52	0.26			

Note. $* p < .05, ** p < .001$

Pairwise comparisons were conducted and indicate that on average, frustration was significantly lower in the no interruption condition when compared to the fine breakpoint condition. However, frustration did not significantly differ between the coarse breakpoint condition when compared to the fine breakpoint condition, or when compared to the no interruption condition.
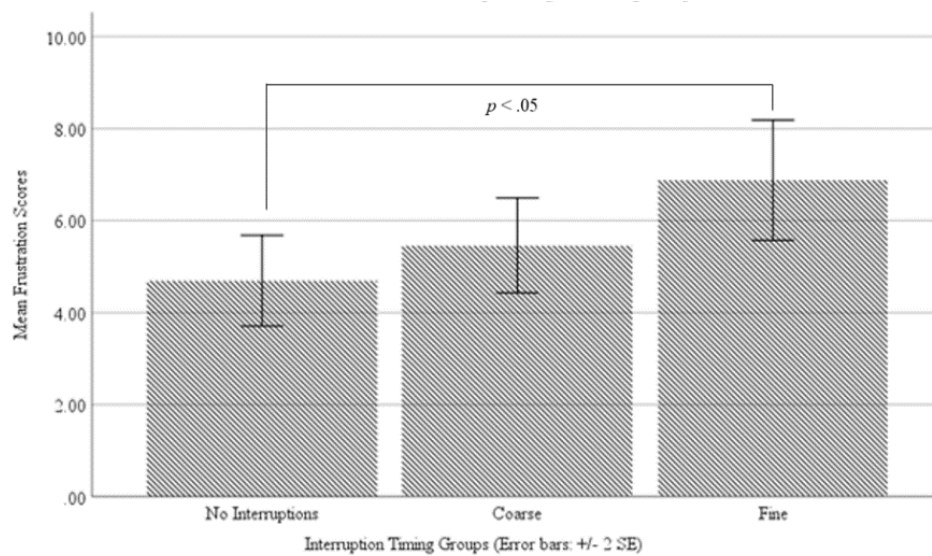
Figure 3. Mean frustration scores within interruption timing conditions. Error bars represent standard errors.

Additional NASA-TLX Analyses

In addition to analyzing total mental workload scores and the frustration subscale, the remaining five subscales were analyzed: mental demand, physical demand, temporal demand, performance, and effort. To assess the differences in each subscale due to interruption frequency and timing, five individual 2 (few, many) X 3 (coarse, fine, none) split-plot ANOVAs were conducted. Based on a Bonferroni correction, statistical significance was assessed at the .008 significance level for all additional NASA-TLX analyses.

The results revealed just two differences. First, physical demand ratings were significantly different based on interruption timing (see Table 12). The means are shown in Figure 4. Pairwise comparisons revealed that interruptions at coarse breakpoints were reported as less physically demanding than interruptions at fine breakpoints. However, there were no differences in physical demand when interruptions occurred at coarse breakpoints compared to receiving no interruptions, or when interruptions at fine breakpoints was compared to receiving no interruptions.

Table 12

Results of the Analysis of Variance for Physical Demand Scores

	SS	df	MS	F	p	partial η^2
Interruption Timing	98.23	2	49.11	7.00	*.001	.119
Interruption Timing x Interruption Frequency	7.26	2	3.63	0.52	.597	.010
Error	729.35	104	7.01			
Interruption Frequency	25.28	1	25.28	0.39	.537	.007
Error	3398.64	52	65.36			

Note. $* p < .05, ** p < .001$
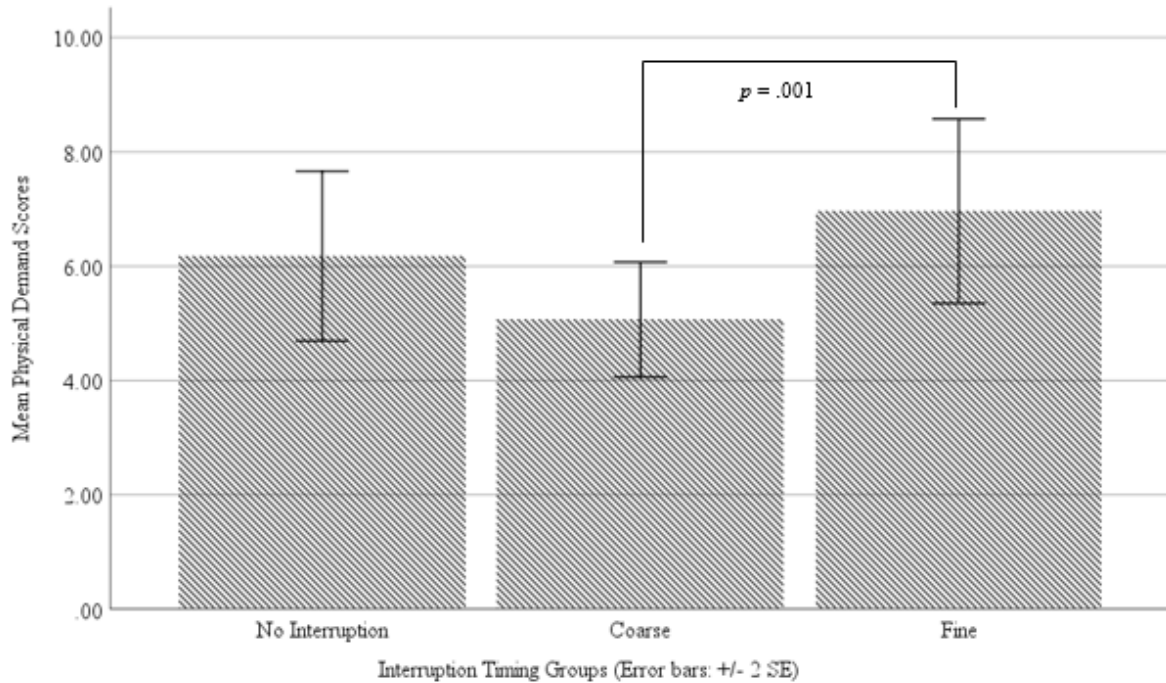
Figure 4. Mean physical demand scores within interruption timing conditions. Error bars represent standard errors.

The results also revealed that performance was reported as significantly better when there were few interruptions compared to many interruptions (see Table 14). The means are shown in Table 13. Note that higher scores indicate better performance. No other differences were found among the remaining subscales.

Table 13

Mean Performance Score within Interruption Frequency Groups and Interruption Timing Conditions

Interruption Timing Groups	Interruption Frequency Groups	n	M	SD
Coarse Breakpoint	Few	27	8.46	3.96
	Many	27	6.76	2.94
	Total	54	7.61	3.56
Fine Breakpoint	Few	27	9.06	4.52
	Many	27	5.83	3.36
	Total	54	7.44	4.27
No Interruption	Total	54	6.97	3.91

Table 14

Results of the Analysis of Variance for Performance Scores

	SS	df	MS	F	p	partial η^2
Interruption Number	11.86	2	5.93	0.97	.382	.018
Interruption Timing x Interruption Frequency	16.01	2	8.01	1.31	.274	.025
Error	634.63	104	6.10			
Interruption Frequency	231.13	1	231.13	7.68	*.008	.129
Error	1565.11	52	30.10			

Note. $* p < .05$

Interruption Frequency

To further explore the nonsignificant findings of interruption frequency, the mean time spent attending to each interruption was analyzed. This was done to understand whether participants spent an equal amount of time attending to the first interruption compared to subsequent interruptions. It was speculated that participants might decrease the time they spent attending to each successive interruption after realizing that the interruptions required no additional action on their behalf and could be disregarded. To assess if there was a difference in average time spent attending to each interruption, two separate analyses were conducted. A one-way ANOVA was conducted to test for differences in the many interruption frequency condition and means were compared with a Bonferroni correction at an adjusted alpha of .008. A t-test was conducted to test for differences in the few interruption frequency condition.

The ANOVA for the many interruption frequency condition revealed that the time spent attending to each interruption was significantly different (see Table 15). Recall that in the many interruptions condition participants received a total of six interruptions, three of which occurred during one task trial (e.g., coarse breakpoints) and another three which occurred during the other task trial (e.g., fine breakpoints). Therefore, the time spent attending to each interruption was analyzed by the order in which participants received the interruptions. Importantly, the fourth interruption in the sequence indicates the start of a new task trial. The means are shown in Figure 5. Pairwise comparisons revealed that time spent attending to the first interruption was significantly longer than the third, fourth, fifth, and sixth interruptions. The second interruption was significantly longer than the third, fifth, and sixth interruptions. The fourth interruption was significantly longer than the sixth interruption.

Table 15

Results of the Analysis of Variance for Time Spent Attending to Interruptions in the Many Interruption Condition

	SS	df	MS	F	p	partial η^2
Interruption Number	609631974.12	5	121926394.83	10.55	**.001	.289
Error	1501900803.88	130	11553083.11			

Note. $* p < .05, ** p < .001$
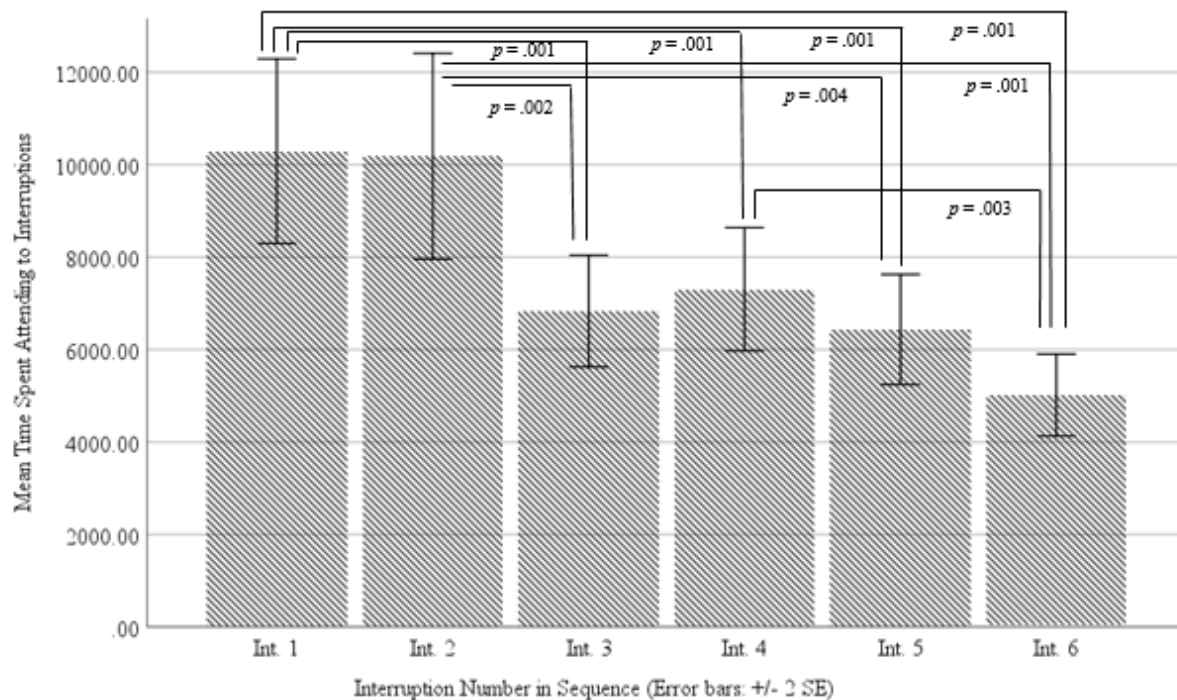
Figure 5. Mean time in milliseconds spent attending to interruptions over time in the many interruption frequency condition. Error bars represent standard errors. Int = interruption.

Recall that in the few interruptions condition participants received only two interruptions, one of which occurred at a coarse breakpoint during one of the task trials and another which

occurred at a fine breakpoint during another task trial. The analysis of the few interruption frequency condition revealed that the time spent attending to the first interruption ($M = 10.39$ sec., $SD = 4.56$ sec.) was significantly longer than the time spent attending to the second interruption ($M = 8.04$ sec., $SD = 4.62$ sec.); $t(26) = 2.57$, $p = .016$.

Baseline Task Performance

To assess if there were differences in baseline performance on any of the task types (restaurant, itinerary, and rental car), performance in the no interruption condition on each task trial was compared. Specifically, four one-way ANOVAs were conducted to assess differences in number of errors, task completion time, total mental workload, and frustration. No analysis was performed for resumption lag because there were no resumption lags in the no interruption condition. A Bonferroni corrected alpha equal to .016 was used for all baseline task comparisons.

The analysis revealed no difference in the number of errors, total mental workload, or frustration between task types. However, task completion time was significantly different between task types. The means are shown in Figure 6. Pairwise comparisons revealed that the restaurant task took significantly longer to complete than the rental car task. However, there was no significant difference in task completion times between the restaurant and itinerary task or the itinerary and the rental car task (see Table 16).

Table 16

Results of the Analysis of Variance for Number of Errors, Task Completion Time, Total Mental Workload, and Frustration

Variable	SS	df	MS	*F*	*p*	partial η^2
Number of Errors						
Task Type	2.93	2	1.46	3.12	.053	.109
Error	23.89	51	0.47			
Task Completion Time						
Task Type	307356.41	2	153678.20	5.59	*.006	.180
Error	140552.34	51	27481.42			
Total Mental Workload						
Task Type	1053.82	2	526.91	1.93	.156	.070
Error	13944.72	51	273.43			
Frustration						
Task Type	0.09	2	0.04	0.29	.748	.011
Error	7.62	51	0.15			

Note. * *p* < .05
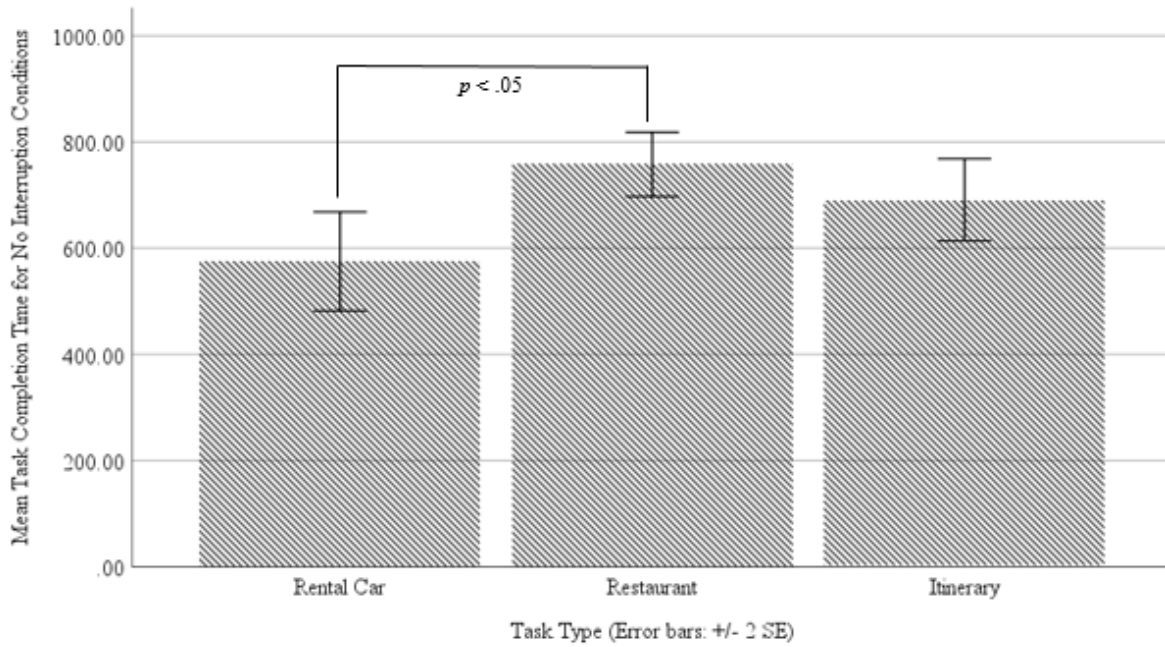
Figure 6. Mean task completion times in seconds for no interruption condition by task type.

Error bars represent standard errors.

CHAPTER IV

DISCUSSION

The purpose of this study was to investigate how event segmentation theory (EST) and hierarchical task analysis (HTA) could be used to determine optimal moments for an interruption. This was accomplished by comparing performance on an itinerary planning task when interruptions occurred at coarse or fine breakpoints. Moreover, this study sought to further our understanding about interruption frequency by comparing performance on the itinerary task when few or many interruptions occurred.

Task Performance

Interruption Timing. One goal of the present study was to investigate why the timing of an interruption impacts its disruptiveness. Previous research has shown that when interruptions occur at inopportune moments, they can negatively impact performance, while interruptions that occur at opportune times can enhance performance (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006). However, identifying where these opportune moments exist within a task has traditionally been a challenge for researchers, especially given that current methods lack a theoretical foundation that would help explain why interruption disruptiveness is influenced by interruption timing. Therefore, this study explored how event segmentation theory (EST) developed by Zacks, Speer, Swallow, Braver, and Reynolds (2007) could be used to determine optimal moments for interrupting a task. Further, this study used hierarchical task analysis (HTA) to identify coarse and fine event boundaries while making predictions in accordance with memory for goals theory (MFG; Trafton et al., 2013).

Number of Errors. As noted above, when an individual works on a task they form a mental representation of that task comprised of different levels, where the highest level represents the overall goal of the task and each level below it represents a division of that goal into subgoals (Iqbal et al., 2005; Salvucci, 2010; Salvucci, Taatgen, & Borst, 2009). The lowest level within the hierarchy is encompasses all the of levels above it. According to Event Segmentation Theory (EST) the point between two subgoals is known as a subtask boundary, and boundaries at higher levels within the hierarchy correspond to coarse breakpoints while boundaries at lower levels correspond to fine breakpoints. Thus, the task representation at a fine breakpoint is broader than at a coarse breakpoint and consequently puts a greater demand on memory to maintain this representation. Based on this greater memory demand, MFG theory suggests that there is an increased chance of forgetting part of the task representation when interrupted at a fine breakpoint. Given these presuppositions, it was predicted that individuals would make more errors when interrupted at fine breakpoints compared to coarse breakpoints.

The results revealed no significant differences in the number of errors made when interrupted at coarse breakpoints, fine breakpoints, or none at all. One reason for the lack of an effect of errors may be due to the few errors that were made ($M = 0.98$). A possible explanation for the absence of an effect may be that the selected breakpoints did not accurately correspond to coarse and fine breakpoints as defined by EST. This topic will be further addressed in the limitations section.

Another possible explanation has to do with the relevance of the interrupting task to the primary task. According to Salvucci, Taatgen, and Borst (2009), if an interrupting task is related to the primary task, recalling the mental representation of the primary task is facilitated by the interrupting task. This occurs because the related interrupting task spreads activation of

information related to the primary task during memory retrieval, decreasing the likelihood of making an error. Recall that Bailey and Konstan (2006) found a difference in the number of errors committed at coarse vs. fine breakpoints. These investigators used interrupting tasks that were unrelated to the primary task and were therefore likely to impede rather than facilitate retrieval. In the present study, however, interruptions were related to the primary task, but did not require the participants to take any additional action. Therefore, the similarity between the interruption(s) and the primary task likely facilitated retrieval and decreased the chance of making an error.

Resumption Lag. According to MFG theory, memory demand is proportional to the time it takes to resume on the primary task after an interruption (Salvucci, Taatgen, & Borst, 2009), meaning that increasing the amount of information stored in memory prior to an interruption also increases the time it takes to retrieve that information and resume the primary task. As stated previously, greater memory demand is needed at fine breakpoints to maintain the broader task representation held in memory. Consequently, it was predicted that interruptions at fine breakpoints would result in significantly longer resumption times than interruptions at coarse breakpoints. The results supported this hypothesis, indicating that on average, participants took 5.4 seconds longer to resume the primary task when interruptions occurred at fine compared to coarse breakpoints, with partial $\eta^2 = .426$ indicating a large effect. Not only does this result replicate previous empirical findings (Adamczyk & Bailey, 2004; Bailey & Iqbal, 2005; Salvucci, 2010), but it expands on those findings by achieving the same result when breakpoints were determined objectively with HTA. Importantly, by using the HTA process, the breakpoints that were generated affected resumption lag as other researchers predicted it would. Additionally, this objective method for determining breakpoints is more robust than subjective methods used

in previous studies because subjective methods can be easily influenced by individual or environmental differences (Graziano et al., 1988; Iqbal et al., 2005; Zacks & Tversky, 2001).

Task Completion Time. MFG theory also suggests that after an interruption at a fine breakpoint an individual must re-encode their task environment (Trafton et al., 2013), which extends the time it takes to complete a task (Salvucci, Taatgen, & Borst, 2009). Conversely, re-encoding is not necessary at coarse breakpoints and therefore does not impact task completion time (Salvucci, 2010). Therefore, it was predicted that task completion time would be longer when an interruption occurred at a fine compared to a coarse breakpoint. Although this effect was not statistically significant, the trend was in the expected direction. An evaluation of the average task completion times for each condition indicated that individuals who were interrupted at fine breakpoints had the longest task completion times, followed by those who were uninterrupted, and finally those who were interrupted at coarse breakpoints. Further, an examination of the mean difference in task completion time between conditions revealed that the difference between the coarse and no interruption condition was smaller than the difference between the fine and no interruption condition, which indicates that the effect of interruptions at coarse breakpoints was most comparable to the no interruption condition. This pattern of results is consistent with that of Bailey and Konstan (2006), who found task completion times that were marginally longer when interrupted at fine as opposed to coarse breakpoints. These researchers also found that when an interruption occurred during completion of a task, users completed that task slower; yet when that same interruption occurred at a coarse breakpoint within the task, there was no effect on completion time.

To understand why interruption timing did not have a significant effect on task completion time, it is important to consider the relationship between re-encoding and errors. Re-

encoding is closely tied to the number of errors made such that when an individual forgets

information related to the primary task (an error), they must retrieve that information by re-

encoding the task environment (Salvucci, 2010). Therefore, because there were no significant

differences in the number of errors made between groups in the present study, it follows that this

would not generate differences in re-encoding and thus in task completion time. Supporting this

notion, factors that were suggested as possible explanations for the lack of significant differences

in errors may also explain the absence of effects on task completion time. Specifically, task

complexity and relevance of the interruption to the primary task are possible factors that

contributed to the lack of differences in errors between groups, while task complexity and

similarity between the primary and interrupting task were found to impact task completion time

(Gillie & Broadbent, 1989). This explanation is also consistent with Salvucci's (2010) statement

that task complexity impacts re-encoding such that basic tasks require little to no re-encoding,

while more complex tasks require several steps of re-encoding. Thus, it is plausible that in the

present study re-encoding was not needed due to the lack of errors made on the itinerary task.

Additionally, Salvucci, Taatgen, and Borst (2009) stated that relevant interruptions may

facilitate retrieval by spreading activation among similar concepts, which consequently allows

for faster and more accurate retrieval of task related information after the interruption. According

to Anderson (1983), this occurs because a task interruption causes the activation associated with

that task to decay. Yet when the interruption is related to the task, it activates a similar network

of activity, keeping some of the primary task active and thus preventing decay. In the present

study, one of the interrupting messages during the itinerary task informed participants that there

was a discount on tickets for certain days of the week. Therefore, the information presented

during the interruption may have activated similar concepts that related to the primary task, such

as types of activities included in the itinerary, hours of operation, time and day of travel, admission prices. Furthermore, this activation may have resulted in fewer errors on the primary task, which consequently reduced the chance of needing to re-encode the task environment.

Mental Workload. In line with the workload aligned task model (WLATM), Bailey and Iqbal (2005) introduced interruptions at the periods of the lowest mental workload within a task, which were determined by observing physiological changes in participant's pupil size. However, the timing of the interruption subsequently did not influence participants self-reports of mental workload, indicating that objective and subjective reports of mental workload did not correspond. Based on this inconsistency in workload measures, responses on the NASA-TLX (Hart & Staveland, 1988) were compared to performance measures. Given that fine breakpoints require a larger task representation to be held in memory, more resources are needed to maintain this task representation, increasing mental workload (Kahneman, 1973). Therefore, it was predicted that subjective mental workload would be significantly higher when interruptions occurred at fine breakpoints than coarse breakpoints.

Consistent with this prediction, the results revealed a significant difference in mental workload with workload significantly higher at fine breakpoints compared to coarse breakpoints with partial $\eta^2 = .179$, indicating a medium effect. Importantly, mental workload was not statistically different when there was no interruption compared to when an interruption occurred at a coarse breakpoint, suggesting that interruptions at coarse breakpoints do not impose a significant cost. This evidence corroborates Bailey and Konstan's (2006) finding that the negative effect of an interruption can be mitigated by executing it at a coarse breakpoint.

Frustration. Finally, based on results of previous research (Adamczyk & Bailey, 2004) it was predicted that individuals would report higher levels of frustration when interrupted at fine

breakpoints than when interrupted at coarse breakpoints. The results were consistent with this research, but the effect was not very strong (partial $\eta^2 = .102$). Although the amount of frustration reported by participants was statistically different between interruption timing conditions, pairwise comparisons revealed that frustration was not significantly different between the coarse and fine breakpoint condition. However, the trend was consistent with predictions. Mean frustration scores revealed that frustration was the lowest when participants received no interruption, followed by when interruptions occurred at coarse breakpoints, and then when interruptions occurred at fine breakpoints.

One potential explanation for this weak effect may be that the order in which participants completed the interruption timing conditions impacted frustration levels. Specifically, it is possible that participants who were exposed to the uninterrupted condition first became more frustrated when the following task trial was interrupted, even if that interruption occurred at a coarse breakpoint. However, upon further examination of frustration levels and the order of the interruption timing conditions, it became clear that there was no order effect.

A second possible explanation for the significant but weak effect of frustration is the difference in methodology between past research that found an effect of interruption timing on frustration and the present study. Predictions were based on the results collected by Adamczyk and Bailey (2004), who found a significant difference in frustration depending on whether an interruption occurred at a coarse breakpoint, fine breakpoint, or not at all. However, this research relied on subjective measures of workload to determine the breakpoints within a task, rather than the more objective approach employed in this study. Thus, the methodology of the current study was different from Adamczyk and Bailey's study and potentially contributed to the inconsistent effect sizes between studies. Additionally, in the Adamczyk and Bailey (2004) study, the

researchers relied on a modified version of the NASA-TLX which truncated participants responses to a scale from 0 to 5 rather than the original scale from 1 to 20, which was used in the present study. By constraining participants' responses, the researchers may have also inflated the differences in frustration levels between conditions.

Interruption Frequency. It was predicted that task performance would be poorer in the high interruption frequency group compared to the low interruption frequency group. Although the performance subscale of the NASA-TLX revealed that participants rated their performance as poorer when many interruptions occurred compared to few interruptions, this rating did not translate to any differences in actual task performance and this hypothesis was not supported.

A possible explanation for these nonsignificant findings is that the nature of the interruptions used in the previous studies differed from that of the current study. Specifically, the interruptions used in the present study were related to the primary task but did not require any additional action to be taken based on the content of the message. For example, while participants completed the itinerary portion of the task, they were presented with a message such as, "Admission prices are 20 percent off when you visit Monday-Thursday". This interruption was related to the task, however, because the trip occurred on a weekend, it did not require participants to alter the itinerary to accommodate the message. Conversely, interruptions used in prior studies required participants to complete some action, such as track a moving target (Monk, 2004) or responding to a question (Lee & Duffy, 2015). Given that the interruption(s) used in the present study did not require additional action, it is plausible that the participants learned that they did not need to pay attention to them and quickly disregarded them.

Recall that all interruptions were approximately 16 words in length and intended to require a comparable amount of time to read. An analysis of the data aimed at the order of

interruptions, however, revealed that the mean time spent attending to each interruption did indeed differ depending on the order. The mean time spent attending to each interruption showed that participants spent the longest time attending to the first interruption, while the time spent attending to successive interruptions grew shorter. The time spent attending to the fourth interruption (i.e., the first interruption in a new task trial) increased slightly, though not statistically, from the time they spent attending to the third interruption (i.e., the last interruption in the previous task trial). Based on this observation, it is likely that participants allocated more time to evaluate the relevance of the first interruption in the new task trial but quickly realized that, similar to the previous task trial, no action was needed and thus decreased the time they spent attending to the remaining interruptions. The post hoc comparisons revealed that for both task trials, participants spent significantly more time attending to the first interruption compared to the last interruption (i.e., interruption one compared to interruption three for the first task trial; interruption four compared to interruption six for the subsequent task trial). This pattern of results was also found in the few interruption condition where participants spent significantly more time attending to the first compared to the last interruption. Based on these findings, there is evidence to suggest that participants learned to disregard interruptions over time and thus minimized the effect of interruption frequency on any of the dependent variables.

Interaction of Timing and Frequency. It was also hypothesized that interruption timing and frequency would interact such that increasing the frequency of interruptions would lead to greater increases in task completion time, number of errors, mental workload, and frustration when interruptions occurred at fine breakpoints compared to coarse breakpoints or not at all. This hypothesis was not supported. As previously discussed, it is likely that the participants realized the interruptions were related to the task but required no further action and thus began to

dismiss the interruptions over time. Consequently, when and how often the interruptions occurred had no impact on performance and therefore there was no interaction between interruption timing and frequency.

Theoretical Implications

The present study applied event segmentation theory (EST) (Zacks et al., 2007) and memory for goals (MFG) theory (Trafton et al., 2013) to understand how interruptions executed at various breakpoints within a task impact performance. Although EST states that interrupting users at different breakpoints within a task will impact disruptiveness, this theory lacks a theoretical foundation to explain why this effect is observed. Therefore, the present study also applied MFG theory to better understand why manipulating the timing of an interruption impacts its disruptiveness.

Most of the findings were in line with event segmentation theory (EST), which proposed that the best time to interrupt an individual is at a coarse breakpoint (Zacks et al., 2007). According to this theory, coarse segments within a task are more salient, better recalled, and better understood than fine segments, making them optimal moments for an interruption. However, EST provides no explanation as to why interruptions at coarse breakpoints should be less disruptive, with some researchers speculating that those boundaries represent moments of lower workload (Bailey & Iqbal, 2008; Bailey & Konstan, 2006; Iqbal et al., 2005; Miyata & Norman, 1986). The present study corroborated these predictions contributed to EST by successfully identifying that mental workload was indeed lower at coarse compared to fine breakpoints. Also, in line with EST's prediction that interruptions at coarse breakpoints are less disruptive, the present study found that resumption lags were shorter at coarse breakpoints in addition to lower frustration.

Additionally, this was the first study to take an objective approach to identify optimal moments for an interruption by utilizing HTA. Previous methods relied on individual and subjective approaches to determine optimal moments, which were often inconsistent across participants. By contrast, in this study the itinerary task was decomposed into subtasks with HTA and the subtasks were them aligned with breakpoints in accordance with EST. Based on the finding that interruptions at fine breakpoints resulted in longer resumption lags, higher mental workload, and frustration levels compared to interruptions at coarse breakpoints, there is support for the success of using HTA as an objective method to determine optimal interruption timing. Importantly, this method can be easily applied to other tasks to determine which moments within the task would be the least susceptible to disruption from an interruption.

These finding are also consistent with memory for goals theory (MFG), which explains how an individual's goals for a task are remembered, retrieved, and forgotten (Trafton et al., 2013). According to this theory, when an interruption occurs during completion of a task, the goals associated with that task become difficult to remember due to decay. The amount of decay depends on a variety of factors, such as the size of the task representation that is held in memory (Salvucci, Taatgen, & Borst, 2009). Specifically, the task representation held an individual's mind at fine breakpoints is larger than at coarse breakpoints and therefore was predicted to result in longer resumption lags, higher mental workload, and frustration after an interruption occurs. As previously stated, the present study supported these predictions; i.e., interruptions at fine breakpoints resulted in longer resumption lags and higher mental workload and frustration when compared to interruptions at coarse breakpoints.

Furthermore, the findings of the present study are consistent with previous research which found that initiating interruptions at coarse breakpoints minimizes their disruptive effects

to a level akin to not being interrupted (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006). Specifically, the present study showed that mental workload and frustration were not significantly different when interruptions occurred at coarse breakpoints compared to conditions when no interruption occurred. Thus, receiving an interruption at a coarse breakpoint was similar to receiving no interruption at all.

Practical Implications

Much of the past research about interruptions has revealed that they have a negative effect on individuals and their task performance (Grundgeiger & Sanderson, 2009; Flynn et al., 1999; Iqbal & Bailey, 2005; Li et al., 2011; Spira & Feintuch, 2005). On the contrary, other research has found that interruptions can be beneficial during a task (Hayes & Miller, 2010; O'Conaill & Frohlich, 1995; Spira & Feintuch, 2005). Consequently, it has been suggested that the timing of an interruption may be responsible for the discrepancy in these findings (Chen, Hart, & Vertegaal, 2007; Dorneich et al., 2012). The findings of the present study supported this idea by showing that interruption timing did in fact influence resumption lag, mental workload, and frustration. Additionally, this effect persisted even when the interruption was something as innocuous as an ill-timed pop-up during a basic planning task. The implication of these findings is that designers should consider the impact of including pop-ups, such as advertisements, as they can detract from the primary task the users were completing prior to being interrupted.

A second important finding of the present study was that interruptions were significantly less disruptive at coarse breakpoints than fine breakpoints, resulting in shorter resumption lags as well as decreased mental workload and frustration. Furthermore, there was no difference in mental workload or frustration when interruptions occurred at coarse breakpoints compared to no interruptions at all. This finding suggests that if a system or individual can decide when to

execute an interruption, doing so at a coarse breakpoint within a task may mitigate, at least some of the negative effects associated with interruptions.

Limitations and Future Directions

There were several limitations in the present study that need to be considered. One potential limitation is that the selected breakpoints did not correspond to participants' individual perceptions of coarse and fine moments. In previous studies of interruption timing, the researchers determined breakpoints by having a group of pilot participants complete a task while reporting where they thought one segment ended and another began (Graziano et al., 1988; Newston, 1973; Zachs et al., 2001). However, because the task was independently decomposed in the present study, it is possible that the task decomposition was not universal for all individuals who completed the task. Consequently, to account for this possibility and ensure the most accurate decomposition, the hierarchical task analysis technique was used. This technique has been successfully used throughout a variety of contexts to systematically identify the subtasks required to complete a task (Annett, 2003). Additionally, in the present study, each task trial was decomposed using this method on two individual occasions to ensure reliability in the resulting hierarchy. Any inconsistency between the hierarchies was remedied by decomposing the task a third time and selecting the hierarchy that was consistent on at least two of the occasions. Moreover, the differences in subjective reports of mental workload suggest that the selected breakpoints corresponded to the coarse and fine breakpoints within the task, further corroborating the approach used.

Another limitation is the potential lack of experience participants had with travel itineraries. For example, given the mean age of participants (M = 20.3) and that many companies require drivers to be 25 years of age to rent a vehicle, it is likely that many participants had no

experience renting a car prior to this experiment. However, an additional analysis the three tasks showed no significant differences in number of errors, mental workload, or frustration between the rental car task trial and the other task trials in the no interruption condition. In fact, the rental car task was completed faster than the other two task types. These results indicate that although participants may have been unfamiliar with renting a car, their baseline performance was similar to the other task trials with which participants were more familiar. However, future work should examine whether the present pattern of results is consistent when other types of tasks are used.

A third limitation is that the number of interruptions in the many interruption condition may have been too few to result in a significant effect. In the many interruption condition, interruption frequency was limited to three interruptions. This value was chosen because past research has shown that when more than three interruptions occur during a task, participants experience high levels of annoyance and frustration, which can cause participants to just give up on the task (Lee & Duffy, 2015; Spier, Valacich, & Vessey, 1999; Xia & Sudharshan, 2002). However, based on the mean frustration levels reported in the present study, it is clear that participants did not experience high levels of frustration in the many interruption condition. Therefore, it is possible that the frequency of interruptions in this condition could have been increased to impact performance without necessarily motivating participants to abandon the task. Based on this limitation, future work should reconsider using three or more interruptions in the many interruption frequency condition.

A fourth limitation is that the interruptions did not require participants to take any action based on the content of the message. Specifically, the interruptions were related to the primary task but required no interaction with the task. This type of interrupting message was used to ensure performance was comparable between the interrupted and uninterrupted conditions.

However, an unintended consequence of relying on interruptions that required no additional action was that participants quickly learned to disregard the interruptions. This was evident based on the decreased time spent attending to interruptions as the number of interruptions increased. Future work should use interruptions that require participants to take some sort of action. For example, an alternative interrupting task that would increase the likelihood of taking action is an instant messaging (IM) conversation (e.g., a friend helping to plan the trip could ask the participant a question requiring an immediate response). By introducing interruptions that require participants to complete some action, they will be forced to attend to each interruption rather than learn to disregard them.

A final limitation is that the primary and interrupting tasks were related to one another and thus potentially facilitated performance on the primary task, thereby eliminating the effect on number of errors and task completion time. According to Salvucci, Taatgen, and Borst (2009), when the interruption and primary task are related, the interruption spreads activation of information that is also related to the primary task during retrieval. Specifically, when an interruption occurs during completion of a task, the activation associated with that task begins to decay (Anderson, 1983). However, when an interruption is related to the primary task, a similar network of activity is activated which ultimately prevents decay by keeping some of the primary task information activated. Consequently, keeping some of the primary task information active helps to reduce the chance of forgetting and subsequently reduces the need to re-encode the task environment. This explanation also accounts for the discrepancy between the present study and past research on the effects of interruption timing on errors and task completion time. Specifically, Bailey and Konstan (2006) found an effect of interruption timing on these variables, however, they used an interrupting task that was unrelated to the primary task, which impeded

retrieval and consequently led to more errors and longer completion times. Therefore, future work should focus on manipulating the relevance of the interruption so it requires a response but one that is unrelated to the primary task. For example, if an IM chat was used as the interruption, the content of the message could include a discussion of the clothing that the "friend" plans to wear. This interruption would still require the participant to take action (i.e., respond to the friend's IM), however, the content of the message would not be related to information in the primary task and therefore should not facilitate primary task activation.

Conclusion

The present study sought to determine whether the timing of an interruption impacts its degree of disruptiveness. This investigation was driven by inconclusive findings as to whether interruptions hinder or help performance on a task (Flynn et al., 1999; Grundgeiger & Sanderson, 2009; Hayes & Miller, 2010; Iqbal & Bailey, 2005; Li et al., 2011; O'Conaill & Frohlich, 1995; Spira & Feintuch, 2005). Additionally, previous methods for investigating interruption timing rely on participants' subjective decomposition of a task, which are often flawed due to the influence of individual or contextual differences (Graziano et al., 1988; Zacks & Tversky, 2001). Therefore, the present study was the first to use hierarchical task analysis (HTA) to decompose a planning task into coarse and fine segments; and then apply event segmentation theory (EST) to identify the coarse and fine breakpoints. Participants completed the planning task divided into three task trials, with interruptions occurring at coarse breakpoints, fine breakpoints, and not at all.

Overall, the general finding of this study was that interruptions may or may not be disruptive, depending on where within the task hierarchy they occur. Specifically, interruptions at fine breakpoints were more disruptive than interruptions at coarse breakpoints, characterized

by an increase in resumption lag, mental workload, and frustration. Moreover, it was discovered that interruptions at coarse breakpoints had effects that were similar to performing a task without interruptions. These findings are consistent with EST and MFG theory, suggesting that the breadth and depth of the task representation changes at various moments within a task, differentially impacting memory demand and consequently performance. However, more work is needed to determine the impact of interruption frequency on performance, as no effect was found in the present study. The lack of an effect was likely due to participants disregarding the interruptions as they increased in frequency. Therefore, future researchers should use interruptions that require the participant to complete some action, thus preventing them from disregarding the interruptions over time. Overall, these results show that the timing of an interruption can have a significant impact on performance, but further research is needed to determine the relationship between interruption timing and frequency.

REFERENCES

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 271-278. ACM.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51,* 355-365.

Annett, J. (2003). Hierarchical task analysis. In A. Duffy (Ed.), *Handbook of Cognitive Task Design*, (pp. 17-35). Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.

Bailey, B. P., Busbey, C. W., & Iqbal, S. T. (2007). TAPRAV: An interactive analysis tool for exploring workload aligned to models of task execution. *Interacting with Computers*, *19*, 314-329.

Bailey, B. P., & Iqbal, S. T. (2005). *Leveraging changes in mental workload during task execution to mitigate effects of interruption* (Technical Report). University of Illinois at Urbana-Champaign.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *14*, 1-28.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*, 685-708.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in

vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine, 78*, B231-B244.

Brixey, J. J., Robinson, D. J., Johnson, C. W., Johnson, T. R., Turley, J. P., & Zhang, J. (2007). A concept analysis of the phenomenon interruption. *Advances in Nursing Science, 30*, E26-E42.

Cain, B. (2007). *A review of the mental workload literature* (Technical Report). Toronto, Canada: Defence Research and Development.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, N.J.: L. Erlbaum Associates.

Chen, D., Hart, J., & Vertegaal, R. (2007). Towards a physiological model of user interruptability. *Human-Computer Interaction*, 439-451.

Clapp, W. C., & Gazzaley, A. (2012). Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiology of Aging, 33*, 134-148.

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2009). Mechanisms of working memory disruption by external interference. *Cerebral Cortex, 20*, 859-872.

Cooper, G. E., & Harper Jr, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities* (AGARD No. 567). Neuilly-Sur-Seine, France: Advisory group for aerospace research and development.

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering etiquette in the design of an adaptive system. *Journal of Cognitive Engineering and Decision Making*, *6*, 243-265.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload,* (pp. 41-62). Amsterdam: North-Holland.

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*, 1319-1325.

Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems*, *16*, 1-47.

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

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance: Cognitive Processes and Performance* (Vol. 2, pp. 41.1—41.49). New York: Wiley.

Graziano, W. G., Moore, J. S., & Collins, J. E. (1988). Social cognition as segmentation of the stream of behavior. *Developmental Psychology*, *24*, 568-573.

Grundgeiger, T., Liu, D., Sanderson, P., Jenkins, S., & Leane, T. (2008). Effects of interruptions on prospective memory performance in anesthesiology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52,* 808-812.

Grundgeiger, T., & Sanderson, P. (2009). Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics*, *78*, 293-307.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Hayes, C. C., & Miller, C. A. (2010). Etiquette considerations for adaptive systems that interrupt. In C.C. Hayes & C. A. Miller (Eds), *Human-computer etiquette: Cultural expectations and the design implications they place on computers and technology* (pp.289-319). Boca Raton: CRC Press.

Hayes, C., Pande, A., & Miller, C. (2002). *Etiquette in human computer interactions: What does it mean for a computer to be polite? Or who needs polite computers anyway?* (Technical Report No. FS-02-02). California: AAAI Press.

Healey, A. N., Primus, C. P., & Koutantji, M. (2007). Quantifying distraction and interruption in urological surgery. *Quality & Safety in Health Care, 16*, 135-139.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45,* 73-93.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*, 104-110.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 311-320). Portland, Oregon: ACM.

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *Conference on Human Factors in Computing Systems: CHI'05 Extended Abstracts on Human factors in Computing Systems* (pp. 1489-1492). Portland, Oregon: ACM.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp.573-605). Amsterdam, Netherlands: Elsevier Science.

Jett, Q. R., & George, J. M. (2003). Work interrupted: A closer look at the role of interruptions in organizational life. *Academy of Management Review*, *28*, 494-507.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 5-39). Amsterdam: North-Holland.

Kahneman, D. (1973). *Attention and effort* (Prentice-Hall series in experimental psychology). Englewood Cliffs, NJ: Prentice-Hall.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*, 72-79.

Lee, B. C., & Duffy, V. G. (2015). The effects of task interruption on human performance: A study of the systematic classification of human behavior and interruption frequency. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *25*, 137-152.

Li, S. Y., Magrabi, F., & Coiera, E. (2011). A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association*, *19*, 6-12.

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind?: Examining the nature of fragmented work. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 321-330). New York: ACM.

Meshkati, N., & Loewenthal, A. (1988). The effects of individual differences in information processing behavior on experiencing mental workload and perceived task difficulty: A preliminary experimental investigation. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 269-288). Amsterdam: North-Holland

Miyata, Y., & Norman, D. A. (1986). Psychological issues in support of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design*: *New Perspectives on Human-Computer Interaction* (pp. 265-284). Hillsdale, NJ: Lawrence Erlbaum.

Monk, C. A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48,* 295-299.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Very brief interruptions result in resumption cost. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1606). Hillsdale, NJ: Lawrence Erlbaum.

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied*, *15*, 291-306.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*, 28-38.

O'Conaill, B., & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. *Conference Companion on Human Factors in Computing Systems: CHI'95 Companion,* (pp. 262-263). New York: ACM.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*, 737-743.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, *21*, 270-297.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload,* (pp. 185-218). Amsterdam: North-Holland.

Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-specific manner. *Applied Cognitive Psychology*, *29*, 680-690.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*, 61-86.

Salvucci, D. D. (2010). On reconstruction of task context after interruption. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2010* (pp. 89-92). New York: ACM.

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2009* (pp. 1819-1828). New York: ACM.

Solovey, E. T., Lalooses, F., Chauncey, K., Weaver, D., Parasi, M., Scheutz, M., ... & Jacob, R. J. (2011). Sensing cognitive multitasking for a brain-based adaptive user interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2011* (pp. 383-392). Vancouver, Canada: ACM.

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

Spira, J. B., & Feintuch, J. B. (2005). *The cost of not paying attention: How interruptions impact knowledge worker productivity* (Technical Report). New York: Basex.

Tams, S., Thatcher, J., Grover, V., & Pak, R. (2015). Selective attention as a protagonist in contemporary workplace stress: Implications for the interruption age. *Anxiety, Stress, & Coping*, *28*, 663-686.

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello II, F. P., Khemlani, S. S., & Schultz, A. C. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, *2*, 30-55.

Trafton, J. G., Jacobs, A., & Harrison, A. M. (2012). Building and verifying a predictive model of interruption resumption. *Proceedings of the IEEE*, *100*, 648-659.

Trafton, J. G., & Monk, C. A. (2007). Task interruptions. In D. A. Boehm-Davis (Eds.), *Reviews of Human Factors and Ergonomics* (pp. 111-126). Santa Monica, CA. Human Factors and Ergonomics Society.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*, 270-285.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, *35*, 263-281.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 63-101). San Diego, CA: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*, 159-177.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton, FL: Taylor and Francis.

Xia, L., & Sudharshan, D. (2002). Effects of interruptions on consumer online decision processes. *Journal of Consumer Psychology, 12,* 265-280.

Xiao, Y. M., Wang, Z. M., Wang, M. Z., & Lan, Y. J. (2005). The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index (English translation). *Chinese Journal of Industrial Hygiene and Occupational Diseases*, *23*, 178-181.

Yakobi, O. (2018). Determinants of association and dissociation between subjective and objective measures of workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62,* 222-226.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*, 979-1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., ... & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651-655.

Zacks, J. M., & Sargent, J. Q. (2010). Event perception: A theory and its application to clinical neuroscience. *Psychology of Learning and Motivation: Advances in Research and Theory, 53*, 253-299.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*, 273-293.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*, 29-58.

APPENDIX A

INFORMED CONSENT FORM
Old Dominion University

PROJECT TITLE: The Application of Event Segmentation Theory to Understanding Interruptions

INTRODUCTION
The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. The title of the study is The Application of Event Segmentation Theory to Understanding Interruptions. This study will be conducted in the Mills-Godwin Building, Room 126.

RESEARCHERS
Mark W. Scerbo, Ph. D., Responsible Project Investigator, College of Sciences, Psychology
Sarah Powers, B.S., Investigator, College of Sciences, Psychology Department

DESCRIPTION OF RESEARCH STUDY
Planning and decision making are complex cognitive processes that can be challenging to study given that we cannot directly observe what is occurring.

If you decide to participate, then you will be one of approximately 100 undergraduate students involved in a study examining planning and decision making behavior. You will be responsible for completing the planning for a hypothetical weekend trip away. You will be given an instruction sheet detailing what aspects of the trip you must plan. While completing the task you may be presented with a message relevant to the task at hand, which must read and acknowledged. The image of the computer display will video recoded so that performance on this task can be analyzed. After completion of each task on the instruction sheet you will be asked to complete a brief questionnaire that assesses the difficulty of the task. If you say YES, then your participation will last no longer than 60 minutes.

EXCLUSIONARY CRITERIA
You will complete a brief background survey to indicate if you have lived in Chicago in the last 5 years and whether English is your primary language. To the best of your knowledge, you should not have lived near the Chicagoland area and English should be your primary language to participate in this study. In addition, you must have normal or corrected-to-normal vision to participate. If you wear contacts or glasses, you must have these with you when you participate.

RISKS AND BENEFITS
RISKS: If you decide to participate in this study, then you may face a risk of some fatigue from completing web-searching tasks. The researchers have tried to reduce these risks by allowing breaks in between each task. And, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: There are no direct benefits for participating in this study.

COSTS AND PAYMENTS
The researchers want your decision about participating in this study to be voluntary. Yet they recognize that your participation may pose as an inconvenience to your time. If you decide to participate in the study, you will receive 1.5 Psychology department research credits, which may be applied to course requirements or extra credit in certain Psychology courses. Equivalent credits may be obtained in other ways, such as conducting library reports and online surveys. You do not have to participate in this study, or any Psychology Department study, in order to obtain this credit.

NEW INFORMATION
If the researchers find new information during this study that would reasonably change your decision about participating, then they will give it to you.

CONFIDENTIALITY
The researchers will take responsible steps to keep private information, such as questionnaire and laboratory performance and findings confidential. You will not be asked to provide any personal identifiers. Information will be stored in a locked filing cabinet prior to its processing. The video camera will be positioned behind you and record an image of the computer display from over your shoulder. You will not appear in the video and no sound will be recorded. You will not be identifiable in the recorded images.

The results of this study may be used in reports, presentations, and publications; but the researcher will not identify

you. Of course, your records may be subpoenaed by court order or inspected by government bodies with oversight authority.

WITHDRAWAL PRIVILEGE
It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study -- at any time. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation. If at any point during the study you wish to stop, simply tell the researcher and you will not be penalized in any way. Any data that has already been collected will be destroyed and will not be included in the final analysis. Your decision will not affect your relationship with Old Dominion University, or otherwise cause a loss of benefits to which you might otherwise be entitled.

COMPENSATION FOR ILLNESS AND INJURY
If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of harm, injury, or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact Dr. Mark Scerbo the primary investigator at (757) 683-4217, Dr. Tancy Vandecar-Burdin the current IRB chair at 757-683-3802 at Old Dominion University, or the Old Dominion University Office of Research at 757-683-3460 who will be glad to review the matter with you.

VOLUNTARY CONSENT
By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researchers should be able to answer them:

Dr. Mark W. Scerbo, mscerbo@odu.edu, (757) 683-4217
Sarah Powers, email: spowe011@odu.edu.

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. Tancy Vandecar-Burdin, the current IRB chair, at 757-683-3802, or the Old Dominion University Office of Research, at 757-683-3460.

And importantly, by signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

Subject's Printed Name & Signature	**Date**

INVESTIGATOR'S STATEMENT
I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

Investigator's Printed Name & Signature	**Date**

APPENDIX B

PARTICIPANT BACKGROUND QUESTIONNAIRE

Participant #:_____   Group:_____   Date:_____   Time:_____

The purpose of this questionnaire is to obtain background information on the participant that will be used for research purposes only.

1. Age_____

2. Gender_____

      0 = Female

      1 = Male

3. Do you have normal or corrected-to-normal vision?_____

      0 = Yes

      1 = No

4. Have you lived in or around the Chicagoland area in the last 5 years?_____

      0 = Yes

      1 = No

5. Is English your primary language?_____

      0 = Yes

      1 = No

APPENDIX C

OVERALL HTA DOCUMENTS

Selecting a rental car task
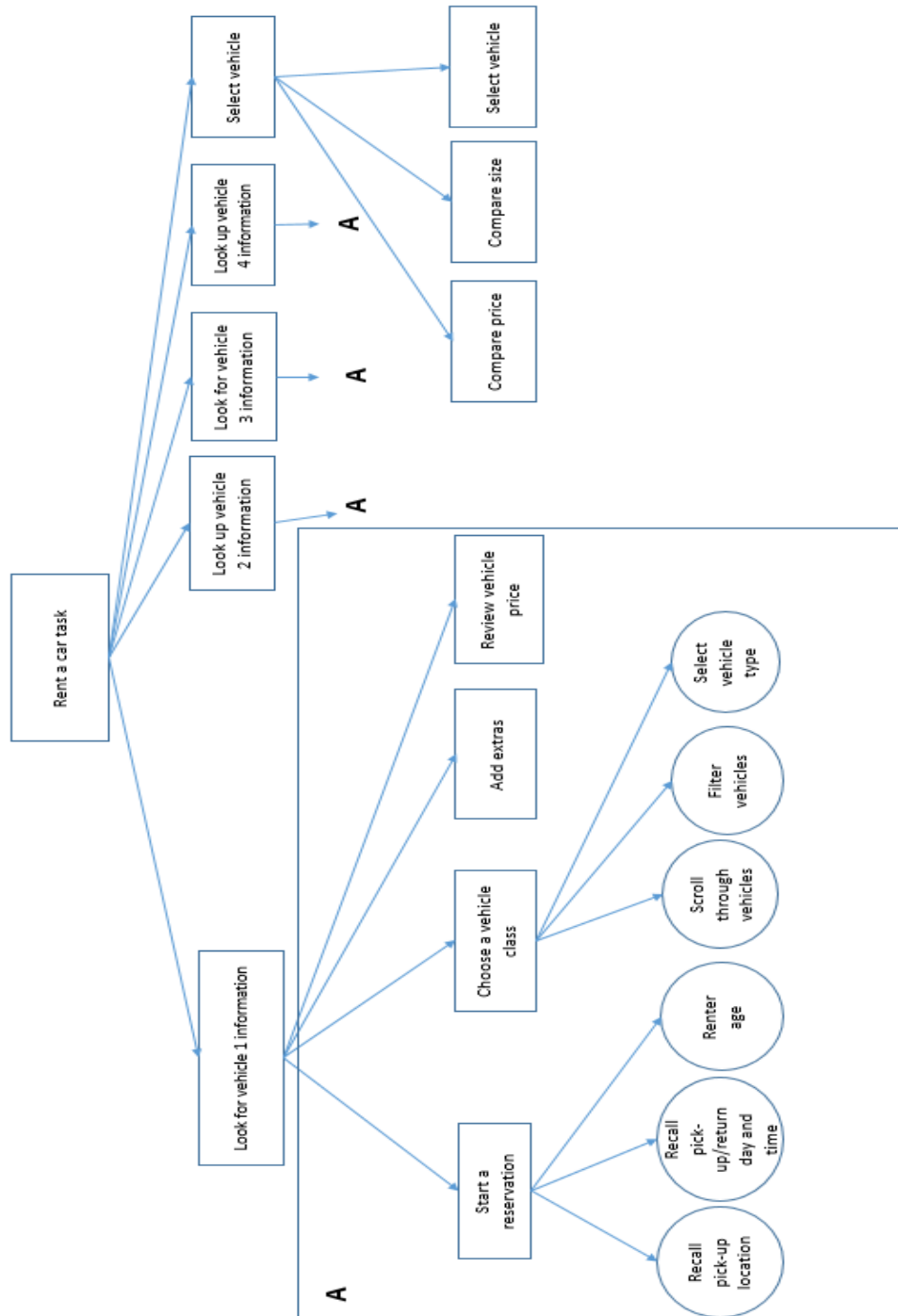
Creating an itinerary task
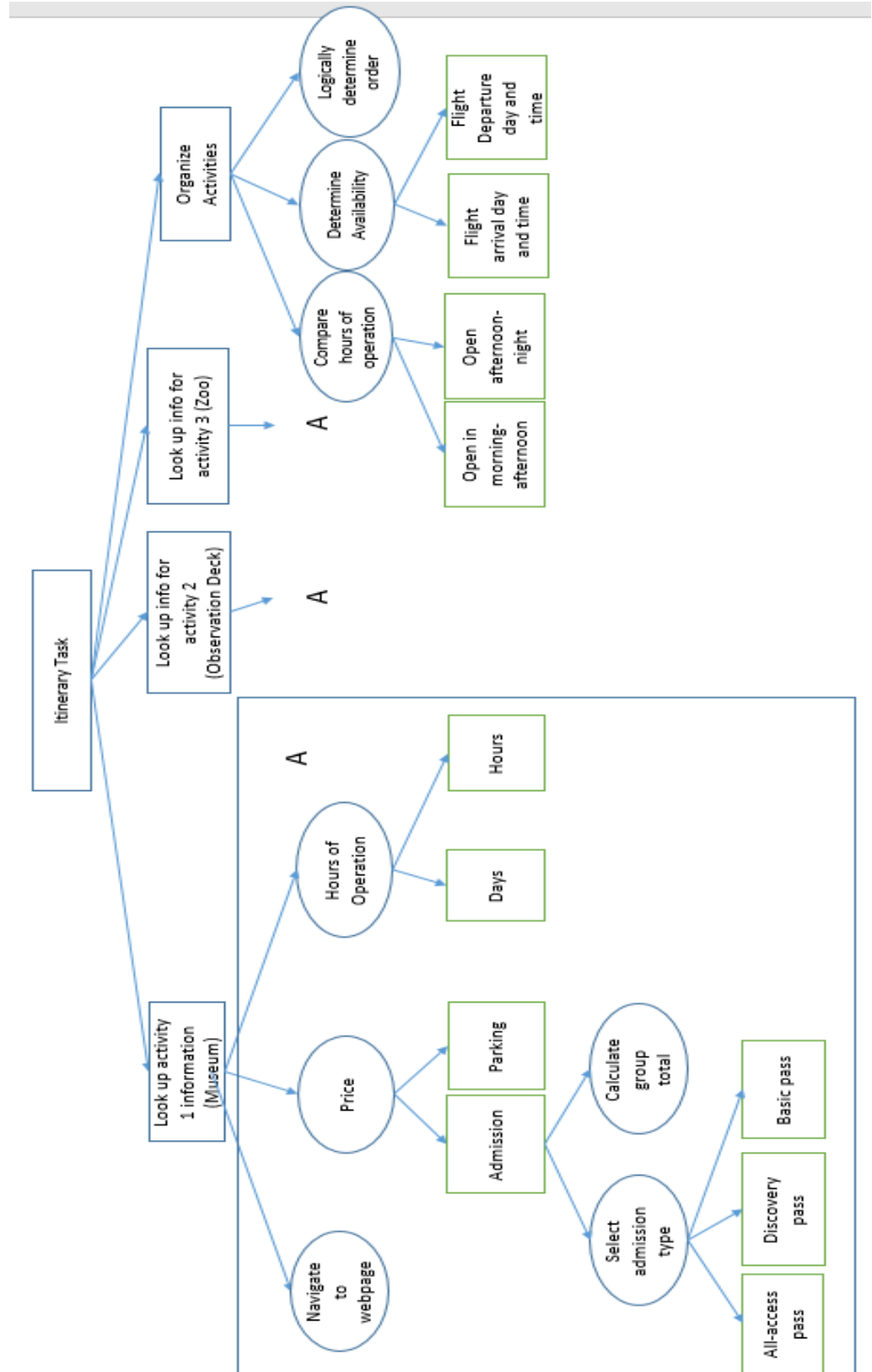
Selecting a restaurant task
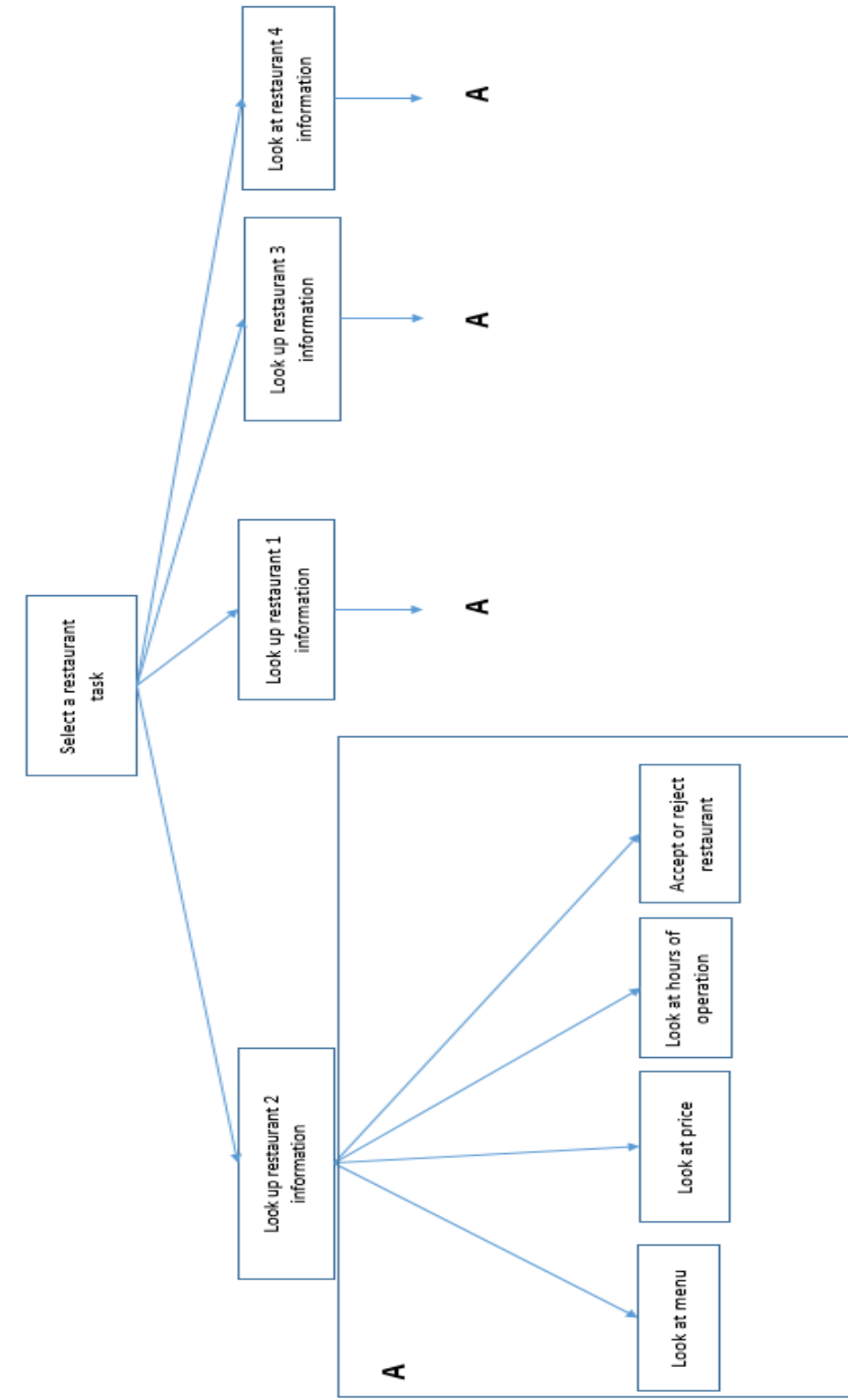
APPENDIX D

DETAILED HTA DOCUMENTS

Selecting a rental car task

0. RENT A CAR

Plan 0: Do a then b or c in any order then d

    a. Navigate to website

    b. Look for vehicle 1 information

    Plan b: Do 1 then 2 then 3 then 4

1. Start a reservation

   Plan 1: Do a, b, c, and d in any order then e

   a. Enter pick-up location

   Plan a: Do i then ii then iii

      i. Recall location

      ii. Type city and/or airport information

      iii. Select "Chicago O'Hare International Airport"

   b. Select pick-up day/time

   Plan b: Do i then ii then iii

      i. Recall information

      ii. Enter pick-up date

         Plan ii: Do 1 then 2

         1. Select month

         2. Select day

      iii. Enter pick-up time

   c. Select return day/time

   Plan c: Do i then ii then iii

      i. Recall information

      ii. Enter return date

         Plan ii: Do 1 then 2

         1. Select month

         2. Select day

      iii. Enter return time

   d. Enter renter age

   Plan d: Do i then ii

      i. Recall Anne's age

      ii. Select "21 years old"

   e. Click continue

2. Choose a vehicle class

   Plan 2: Do a or b then c

   a. Scroll through vehicles

   Plan a: Do i then ii

      i. Recall vehicle type

      ii. Scroll and find "Intermediate SUV"

   b. Filter for vehicles

Plan b: Do i or ii or both then iii

    i. Vehicle type

    ii. Passenger capacity

    iii. Scroll and find "Intermediate SUV"

  c. Select "Intermediate SUV"

3. Add Extras

Plan 3: Do a then b then c

  a. Recall extras for vehicle 1

  b. Add "Sirius Xm Radio"

  c. Click "continue to review"

4. Review vehicle price

c. Look for vehicle 2 information (repeat b but with vehicle 2 information)

d. Look for vehicle 3 information (repeat b but with vehicle 3 information)

e. Look for vehicle 4 information (repeat b but with vehicle 4 information)

f. Select Vehicle

Plan d: Do i or ii or both then iii

  i. Compare prices

  ii. Compare size

  iii. Select better option

Create an itinerary task

0. **PLAN AN ITINERARY**
   Plan 0: Do 1 or 2 or 3 then 4
1. **Look up Activity 1 Information (Brookfield Zoo)**
   Plan 1: Do a then b
   a. Navigate to page
      Plan a: Do i or ii
      i. Google
         Plan i: Do 1 then 2
         1. Type "Brookfield zoo"
         2. Click on Brookfield Zoo page
      ii. Click on link
   b. Look up hour and pricing information
      Plan b: Do i or ii; if you do ii you must also do i
      i. Click on "Plan your visit"
         Plan i: Do 1 or 2
         1. Click on "Hours and prices"
            Plan for 1: Do a then b then c then d
            a. Zoo hours
            b. Admission
               Plan b: Do i then ii then iii
               i. Locate price of adult ticket
               ii. Calculate total group cost
               iii. Record total price
            c. Parking
            d. Extra shows/zoo attractions
         2. Click on "Group Sales"
      ii. Click on "Tickets"
         Plan ii: Do 1 or 2
         1. Click on "Group Sales"
         2. Click on "General Admission"
            Plan 2: Do a then b then c
            a. Click "add to cart"
            b. Type in number of tickets needed
            c. Record total price
2. **Look up Activity 2 information (Field Museum)**
   Plan 2: Do a then b or c or d
   a. Navigate to page
      Plan a: Do i or ii
      i. Google
         Plan i: Do 1 then 2
         1. Type "Field Museum"
         2. Click on Field Museum page

      ii.   Click on link

  b. Click on "Tickets"

      Plan b: Do i and ii in any order

      i.   Select Pass type

         Plan i: Do 1 or 2 or 3 then 4 then 5

         1. All-access pass

         2. Discovery pass

         3. Basic admission

         4. Calculate group total cost

         5. Record total cost

      ii. Look at hours of operation

  c. Click on "Plan your Visit"

      Plan c: Do i and ii in any order

      i.   Select Pass type

         Plan i: Do 1 or 2 or 3 then 4 then 5

         1. All-access pass

         2. Discovery pass

         3. Basic admission

         4. Calculate group total cost

         5. Record total cost

      ii. Look at hours of operation

  d. Click on "Menu"

      Plan d: Do i or ii

      i.   Click on "Tickets"

        Plan i: Do 1 and 2 in any order then 3 then 4

         1. Select pass type

            Plan 1: Do a or b or c

            a. All-access pass

            b. Discovery pass

            c. Basic admission

         2. Look at hours of operation

         3. Calculate group total cost

         4. Record information

      ii. Click on "General Admission"

3. **Look up Activity 3 Information (Observation Deck)**

    Plan 3: Do a then b or c or d and e in any order

  a. Navigate to page

      Plan a: Do i or ii

      i.   Google

        Plan i: Do 1 then 2

         1. Type "Skydeck Chicago"

         2. Click on Skydeck page

      ii. Click on link

  b. Find ticket pricing and hours of operation

Plan b: Do i or ii or iii or iv

   i.   Click on "Plan your visit"

     Plan i: Do 1 then 2 or 3 then 4 then 5

       1.   Click on "hours and pricing"

       2.   Identify hours of operation

       3.   Scroll through admission types

         Plan 2: Do a

          a.   Adult admission

       4.   Calculate group total

       5.   Record information

   ii.   Click on "Buy tickets"

     Plan ii: Do 1

       1.   Go back to main menu

   iii.   Click on "hours and pricing"

     Plan iii: Do 1 or 2 then 3 then 4

       1.   Identify hours of operation

       2.   Scroll through admission types

         Plan 2: Do a

          a.   Adult admission

       3.   Calculate group total

       3. Record information

   iv.   Click on "visit"

     Plan iv: Do 1 then 2 or 3 then 4 then 5

       1.   Click on "hours and pricing"

       2.   Identify hours of operation

       3.   Scroll through admission types

         Plan 2: Do a

          a.   Adult admission

       4.   Calculate group total

       5.   Record information

4. Organize Activities

Plan 4: Do a and b in any order then c

   a.   Compare hours of operation

     Plan a: Do i and ii in any order

       i.   Open in mornings

       ii.   Open in afternoon/night

   b.   Figure out availability

     Plan b: Do i and ii in any order

       i.   Days

         Plan i: Do 1 and 2 in any order

1. Recall arrival day
2. Recall departure day

    ii. Time

Plan ii: Do 1 and 2 in any order

1. Recall arrival time
2. Recall departure time

c. Deduct which activity should occur when

Plan c: Do i and ii and iii and iv in any order

    i. Museum and Zoo cannot be same day
    ii. Museum and Zoo must be in the morning-afternoon
    iii. Skydeck must be in afternoon-night
    iv. Skydeck must be Saturday

Select a restaurant task

0. **SELECT A RESTAURANT**
   Plan 0: Do 1, if X is present stop, if X is not present do 2 and so on
1. **Look up Restaurant 1 Menu (Oriole)**
   Plan 1: Do a then b then c
   a. Navigate to webpage
   b. Click on "Menu"
      Plan b: Do i and ii in any order
         i. Look for vegan options
         ii. Look at price
   c. Determine if restaurant meets group requirements
2. **Look up Restaurant 2 Menu (Native Foods Café)**
   Plan 2: Do a then b and c in any order then d
   a. Navigate to webpage
   b. Click on "Menu"
      Plan b: Do i and ii in any order
         i. Look for vegan options
         ii. Look at price
   c. Click on "Place"
      Plan c: Do i then ii
         i. Select location
         ii. Select carrot to view hours of operation at selected location
   d. Determine if restaurant meets group requirements
3. **Look up Restaurant 3 Menu (Handlebar)**
   Plan 3: Do a then b then c then d
   a. Navigate to webpage
   b. Scroll to hours of operation
   c. Click on/scroll to "Menu"
      Plan b: Do i and ii in any order
         i. Look for vegan options
         ii. Look at price
   d. Determine if restaurant meets group requirements
4. **Look up Restaurant 4 Menu (Osteria Langhe)**
   Plan 4: Do a then b then c
   a. Navigate to webpage
   b. Click on "Menu"
      Plan b: Do i and ii in any order
         i. Look for vegan options
         ii. Look at price
   c. Determine if restaurant meets group requirements

APPENDIX E

TASK INSTRUCTION SHEET- SEE BELOW

Your friend Anne started planning a weekend trip away to Chicago for you, two friends, and Anne. She suddenly got swamped at work and has asked you to pick up where she left off and finish planning the trip and keep a running total of trip expenses.

You will complete three tasks which include selecting a rental car, planning an itinerary, and choosing a restaurant using the information Anne provided. You must complete these tasks without exceeding the $750 budget the group has agreed upon. You can use the budget at your discretion, however once you determine a total for a part of the trip that amount is final and cannot be changed after moving onto other parts of the trip.

Anne has already determined that you will fly into Chicago O'Hare International Airport on Friday August 16th at 9pm and depart from the same airport on Sunday August 18th at 4pm.

Feel free to use the textboxes provided to record any notes or comments.

This page is intentionally left blank. Scroll down when you are told to do so.

Anne decided that you will rent one of four vehicle types from Enterprise. Anne has additionally requested that you calculate the cost of each vehicle with several added features, such as a GPS. The features Anne wants you to add to each vehicle are listed below and can be found on the extras page during the reservation process. You will go through the reservation process four times (once for each vehicle) using the Enterprise webpage and report the total price for each vehicle found on the final page of the reservation process. You will then decide which vehicle is best for the trip by considering things such as price, size, practicality, etc.

Note: Anne will be driving the vehicle so when it asks for the renter's age you can enter Anne's age, 21.

Rental Car: https://www.enterprise.com/en/home.html

Vehicle 1: "7-Passenger Minivan"

Extras: Greenhouse gas emissions offset and Sirius XM radio

Vehicle price:

Vehicle 2: "Standard"

Extras: Greenhouse gas emissions offset and GPS

Vehicle price:

Vehicle 3: "Full Size"

Extras: Sirius XM radio and GPS

Vehicle price:

Vehicle 4: "Intermediate SUV"

Extras: Greenhouse gas emissions offset and Sirius XM radio

Vehicle price:

Vehicle Selected:

Estimated Rental Expense:

Estimated Trip Total:

Your friend Anne started planning a weekend trip away to Chicago for you, two friends, and Anne. She suddenly got swamped at work and has asked you to pick up where she left off and finish planning the trip and keep a running total of trip expenses.

You will complete three tasks which include selecting a rental car, planning an itinerary, and choosing a restaurant using the information Anne provided. You must complete these tasks without exceeding the $750 budget the group has agreed upon. You can use the budget at your discretion, however once you determine a total for a part of the trip that amount is final and cannot be changed after moving onto other parts of the trip.

Anne has already determined that you will fly into Chicago O'Hare International Airport on Friday August 16th at 9pm and depart from the same airport on Sunday August 18th at 4pm.

Feel free to use the textboxes provided to record any notes or comments.

This page is intentionally left blank. Scroll down when you are told to do so.

On the night of your arrival in Chicago, the group wants to eat at a local restaurant. Four restaurants have been suggested; however, Anne did not have time to research each one. It is up to you to find additional information and determine the best restaurant choice for the group. It is important to note that Anne is a vegan and requested that the selected restaurant has several vegan options on the menu. In addition, you should check if the restaurant is open on the night of arrival and within the trip budget. You can calculate the approximate total by looking at menu prices to determine the average price per meal. For each restaurant please research and record if it is open the night of arrival, the estimated group cost, and if there is more than one vegan option even if that restaurant does not meet one of the requirements.

Native Foods: https://www.nativefoods.com/

Open on night of arrival?

Estimated group cost:

More than 1 vegan option?

Oriole: http://www.oriolechicago.com/

Open on night of arrival?

Estimated group cost:

More than 1 vegan option?

Handlebar: https://www.handlebarchicago.com/

Open on night of arrival?

Estimated group cost:

More than 1 vegan option?

Osteria Langhe: https://www.osterialanghe.com/

Open on night of arrival?

Estimated group cost:

More than 1 vegan option?

Restaurant Selected:

Estimated Group Restaurant Expense:

Estimated Trip Total:

Your friend Anne started planning a weekend trip away to Chicago for you, two friends, and Anne. She suddenly got swamped at work and has asked you to pick up where she left off and finish planning the trip and keep a running total of trip expenses.

You will complete three tasks which include selecting a rental car, planning an itinerary, and choosing a restaurant using the information Anne provided. You must complete these tasks without exceeding the $750 budget the group has agreed upon. You can use the budget at your discretion, however once you determine a total for a part of the trip that amount is final and cannot be changed after moving onto other parts of the trip.

Anne has already determined that you will fly into Chicago O'Hare International Airport on Friday August 16th at 9pm and depart from the same airport on Sunday August 18th at 4pm.

Feel free to use the textboxes provided to record any notes or comments.

This page is intentionally left blank. Scroll down when you are told to do so.

Your friends requested that you put together an itinerary to see the Chicago Observation Deck, the Field Museum, and the Brookfield Zoo over the course of the weekend. To do this, you must research each activity to determine the admission price and hours of operation. After researching each activity, use that information to strategically plan the order the group should complete each activity and on which days. Keep in mind that each activity will take a minimum of 5 hours to complete.

Weekend Activities:

Observation Deck: https://360chicago.com/

Admission price:

Hours of operation:

Field Museum: https://www.fieldmuseum.org/

Admission price:

Hours of operation:

Brookfield Zoo: https://www.czs.org/Brookfield-ZOO/Home

Admission price:

Hours of operation:

Use the space on the next page to record the itinerary information (i.e. order of the activities).

Itinerary

First activity:

Visit day/time:
Group admission total:

Second activity:

Visit day/time:
Group admission total:

Third activity:

Visit day/time:
Group admission total:

Estimated Group Activities Expense:

Estimated Trip Total:

APPENDIX F

NASA-TASK LOAD INDEX (TLX) WORKLOAD QUESTIONNAIRE

(Hart & Staveland, 1988)

MENTAL DEMAND

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

PHYSICAL DEMAND

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

TEMPORAL DEMAND

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

PERFORMANCE

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

EFFORT

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

FRUSTRATION

Low   High

| ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- |

APPENDIX G

NASA-TLX SCRIPT

After you complete each task, you will fill out a NASA-TLX survey, which looks like this..(show participants survey). For each task you will report how much mental demand, physical demand, temporal demand, quality of performance, effort, and frustration level you experienced. Each of these can be reported by circling the vertical line, which ranges from low to high, or good to poor for the performance subscale (point out on survey). Each scale ranges from (1); indicating a low or poor score, to (20); indicating a high or good score.

Notice that low is on the left for every subscale, except for performance. The performance subscale trends in the opposite direction, where better performance is indicated by circling closer to the left of the scale and poorer performance is indicated by circling closer to the right of the scale (point out on survey).

Do you have any questions about reporting?

Now that you understand how to report, I will go through each of the subscales and define what they mean. If at any time you would like me to repeat what a subscale means just simply ask and I would be more than happy to explain.

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	good/poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Figure 8 NASA-TLX Rating scale definitions.

Hart, S. G., & Staveland, L. E. (2005). Results of empirical and theoretical research. *Ergonomics: Skill, Displays, Controls, and Mental Workload*, *2*, 408.

Now that we've gone over the definitions for each subscale do you have any questions, concerns, or anything that needs clarification about the NASA-TLX?

APPENDIX H

LIST OF INTERRUPTIONS BY TASK

List of Interruptions by task (~16 words each)

1. Rent a car
   a. Q: Enterprise now provides the option to purchase this vehicle at the low rate of $250/month!
   b. W: All Enterprise vehicles come with a handicap accessible add-on for an extra $25 per day.
   c. E: Unfortunately, the 7-seat minivan is currently unavailable. We apologize for any inconvenience this may cause.
   d. R: Starting today, all 2-door sports car rentals are available with a 5 percent rental fee reduction!
   e. T: For a limited time, receive a free vehicle rental after booking four+ rentals with Enterprise.
   f. Y: When you book a vehicle with Enterprise you are eligible for $50 off select hotel rooms.

2. Plan an itinerary
   a. A: The weekend August 16th-18th, Chicago Skydeck (The Museum, Brookfield Zoo) will be closing an hour early, due to renovations.
   b. S: For a limited time, Seaworld tickets are 50 percent off when you purchase Chicago Skydeck (The Museum, Brookfield Zoo) admission.
   c. D: Tired of the long walk from the parking lot? Valet parking will soon be available!
   d. F: Group tickets (parties of 10 or more) receive an extra 10 percent discount on admission prices!
   e. G: Weekday admission discount! Admission prices are now 20 percent off when you book a visit Monday-Thursday.
   f. H: Now offering a Chicago appreciation discount! Any Chicago resident will receive $5 off admission prices!

3. Pick a restaurant
   a. K: BOGO dinner! Buy one dinner and receive a free dinner within one week of purchase.
   b. X: Diner special! Kids eat free Tuesday through Thursday with the purchase of one adult entrée.
   c. B: Download our new food lover app to submit orders and make reservations with your smartphone!
   d. N: We offer a 10 percent discount for any reservations made at least 1 year in advance.
   e. M: Receive one wine bottle of your choice when you spend more than $250 (before tax).
   f. L: Back to school deal! 5 percent discount for students who show a valid high school ID.

APPENDIX I

DEBRIEFING SCRIPT

Thank you for your time and participation in our research study.

I hope you enjoyed your experience and learned some things today. If you have any questions later, or would like to know the results of this study, please feel free to contact:

Dr. Mark W. Scerbo, mscerbo@odu.edu, (757) 683-4217
Sarah Powers, email: spowe011@odu.edu.

Within the next few days, you will receive one credit for participating in this study. If you have any concerns with receiving this credit, feel free to contact me directly:

Sarah Powers, email: spowe011@odu.edu.

Do you have any other questions or comments about anything you did today or anything we've talked about?

Thank you again for your participation.

APPENDIX J

INTRODUCTION SCRIPT

Welcome, thank you for coming in!

Before we begin, please silence your device. If it goes off during the experiment, I will ask you to leave.

First, here is the informed consent form, I will go over what it says briefly…

Do you have any questions about that?

Second, I need you to complete this brief background questionnaire to determine if you pass all our exclusionary criteria…

Thanks, did you have any questions about that?

Great! So today you will be completing the planning of a hypothetical weekend trip away to Chicago. To plan the trip you will complete three tasks which include creating an itinerary, picking a rental car, and picking a restaurant.

To complete each of these tasks, you will have to use information given in the task instructions and the website links provided to research any information needed to inform your decisions. You can take notes and record your decisions on the spaces provided on the word document.

You cannot skip any sections of the task for any reason. Additionally, you must research the information in the order it is presented to you. For example, if you are given four vehicles to research ordered 1-4, you must start with vehicle 1 as opposed to picking a vehicle from anywhere on the page that would like to research first.

When you are researching you must go back and forth between researching and entering information on the word document. In other words, you cannot have the webpage and the word document side by side. This is to ensure the camera can always see clearly.

Any questions?

Okay we are going to go over the overall weekend trip instructions you see on the screen in front of you. These instructions will help you to complete each of the three tasks….

Any questions about that?

Final important things to know before we get started:

While you are performing the task, you may see pop-up messages with information relevant to the task. Be sure to read those pop-ups and then click the x in the upper right corner to close the message.

After you complete a task, do not scroll to a new section until you are told to do so because after each task you will fill out a survey about the amount of mental workload required to complete that task. You will repeat this process three times, once for each task.

Now I will familiarize you with the NASA-TLX…

Any questions about that?

Aright were going to make sure the GoPro has a good view of the screen.

Before we begin do you have any questions? Ask now because you should try and avoid asking any questions during the experiment.

APPENDIX K

SAMPLE PARTICIPANT CHECKLIST

Example-few interruption condition (order of task trials: rental car, restaurant, itinerary; order of interruption timing: coarse breakpoint, fine breakpoint, none)

RC, C	RR, F	IT, N
1. After recording vehicle 1 price before researching vehicle 2 info **CNTRL W**	1. After reviewing menu prices at RR1 before recording average meal price **CNTRL K**	N/A

Example- many interruption condition (order of task trials: restaurant, itinerary, rental car; order of interruption timing: fine breakpoint, none, coarse breakpoint)

RR, F	IT, N	RC, C
1. After reviewing menu prices at RR1 before recording average meal price **CNTRL K** 2. After reviewing menu prices at RR2 before recording average meal price **CNTRL X** 3. After reviewing menu prices at RR3 before recording average meal price **CNTRL N**	N/A	1. After recording vehicle 1 price before researching vehicle 2 info **CNTRL W** 2. After recording vehicle 2 price before researching vehicle 3 **CNTRL T** 3. After recording vehicle 3 price before researching vehicle **CNTRL Y**

<div align="center">

VITA

Sarah Anastazia Powers

</div>

Department of Psychology
Old Dominion University, Norfolk, VA 23529

EDUCATION

Old Dominion University, Norfolk VA   **2017-Ongoing**
Ph.D. Psychology: Human Factors   *Expected May 2022*
M.S. Psychology   *Expected October 2019*
Iowa State University, Ames IA   **2013- 2016**
B.S. magna cum laude in Psychology
B.A. magna cum laude in Communication Studies

RESEARCH EXPERIENCE

Undergraduate Research Assistant   **Spring 2014**
Department of English, Iowa State University

Research Assistant   **Spring 2016**
Navigation Lab, Iowa State University

Graduate Research Assistant   **Fall 2017- Ongoing**
Simulation Usability Research Facility, Old Dominion University

PUBLICATIONS

Fluegel, B.A., **Powers, S.**, Scerbo, M (2018). The effects of interruptions on novices learning a challenging surgical skill. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA.

PROFESSIONAL EXPERIENCE

Academic Tutor   **Fall 2014- Spring 2015**
Academic Success Center, Iowa State University

Graduate Teaching Assistant   **Fall 2017- Ongoing**
Old Dominion University

PROFESSIONAL MEMBERSHIPS

ODU HFES   **Fall 2017- Ongoing**
Student Member   *Fall 2017-Summer 2018*
Secretary   *Fall 2018- Summer 2019*
Vice President   *Fall 2019-Ongoing*
Psychology Club   **Fall 2015- Spring 2016**
Student Member
Psi Chi   **Fall 2015- Spring 2016**
Student Member
Mortar Board   **Fall 2015- Spring 2016**
Student Member

ADDITIONAL EXPERIENCE & SKILLS

SPSS
Study Abroad   **Summer 2015**
Modeling and Simulation Certificate   **Fall 2018**