Interruption processing in a decision-making task:
Successful integration of interruptions and task resumption

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

By

Sheryl Miller
Bachelor of Arts
State University of New York, College at Geneseo, 1995
Master of Arts
George Mason University, 1997

Director: Deborah A. Boehm-Davis, Professor
Department of Psychology

Spring Semester 2004
George Mason University
Fairfax, VA

DEDICATION

This dissertation is lovingly dedicated to my mother, Dolores Miller, who always fostered in me a love of learning and a sense of self-assurance and determination that served me well in my education and to my husband, Adam Tucker, who showed unfailing patience, consideration, and devotion during my graduate studies especially while I ruminated over every element of this dissertation.

ACKNOWLEDGEMENTS

I am thankful for the support of my dissertation chair, Dr. Deborah Boehm-Davis, and the members of my committee, Dr. Greg Trafton, Dr. Erik Altmann, and Dr. Leonard Adelman. Without their support during my graduate career and in formulating and carrying out this research, this dissertation would not have been written.

I would also like to express my appreciation to those who have actively assisted me in carrying out this research: Dr. Michael Schoelles for masterminding the I-ARGUS task and patiently supporting my programming needs; Dr. Wai-tat Fu and William Pennell for their creative programming support; and Daragh Sibley for his reliable assistance in participant recruitment and data collection.

My work has benefited from conversations with many colleagues, friends, and acquaintances. In particular, I have been delighted to work with and received professional and personal inspiration from faculty and fellow students in the ARCH Lab, especially Melanie Diez, Dr. Wayne Gray, Dr. DeVere Henderson, Audrey Lipps, Christopher Monk, Dr. Susan Trickett, and Cedric Yeo. I especially appreciate the friendly support and practical guidance of Dr. Leonard Adelman that was invaluable to my professional development. I would also like to express appreciation to Dr. Michael Venturino and Dr. Karen Duffy whose professional examples and dedication to undergraduate teaching and research motivated my career choices.

I would also like to express appreciation to several people in particular who offered friendship and collegial empathy during my graduate career, including Colby Mills, Dr. Jill Emanuele, and Dr. Vicki Strnatka-Knapp. Vicki, in particular, has been a wonderful friend and witness to the challenges of my own doctoral process starting with graduate school applications and culminating with this dissertation.

Finally, the writing of a dissertation can be an isolating experience that is not possible to sustain without the practical and emotional support of numerous friends and family. My sincere gratitude goes to my husband, Adam Tucker, my mother, Dolores Miller, and my adopted family, Joanne Tucker and Carrie Tucker, for their steadfast encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

INTERRUPTION PROCESSING IN A DECISION-MAKING TASK: SUCCESSFUL INTEGRATION OF INTERRUPTIONS AND TASK RESUMPTION

Sheryl Miller, M.A.

George Mason University, 2004

Director: Dr. Deborah A. Boehm-Davis

Previous research has ignored interruption contexts where there are competing memory requirements associated with the interruption itself and the resumption of the interrupted task. This complexity was investigated in a simulated team decision-making task where operators needed to remember both the interrupting message and the resumption point of the interrupted decision task. An initial experiment was used to assess the complex effects of different interruptions on initial and subsequent decisions. Interruptions resulted in longer decision times for the interrupted task when messages needed to be remembered over time. However, subsequent decisions were unaffected, and the eventual recall of message data actually resulted in faster decisions due to participant adaptation. A modest strategy to rehearse the task resumption point during the interruption lag was proposed but actually increased interrupted decision time. Participants who tried to remember both the message and the task resumption point performed worse, likely because of interference between the two items. The next three experiments tested this hypothesis by manipulating the costs of forgetting the interrupting message and forgetting the task resumption point. Experiments 2 and 3 differed in the complexity of the decision task. Experiment 4 manipulated the interface so that people had more opportunity to ignore messages. As predicted, participants consistently considered trade-offs between forgetting the interrupting message and forgetting the task resumption point.

1. INTRODUCTION

Interruption is both an adaptive part of communication and a potentially calamitous intrusion. It is this dual perspective that makes interruption an interesting topic for investigation. However, the research literature has not provided a very sophisticated understanding of interruptions as complex and consequential, nor has the literature offered solid explanations of how people process interruptions. Past research has focused on how people resume an interrupted task after attending to some unrelated secondary task, ignoring interruptions that are an integral part of overall performance. In many contexts, such as flightdeck operations (Bainbridge, 1984; Dismukes, Young, & Sumwalt, 1998; Loukopoulos, Dismukes, & Barshi 2001), using in-vehicle devices (McKnight & McKnight, 1993; Stevens & Minton, 2001; Llaneras & Singer, 2003; Patten, Kircher, Ostlund, & Nilsson, 2003), and group collaboration (Dix, Ramduny, & Wilkinson, 1995; Fussel, Kraut, Lerch, Scherlis, McNally, & Cadiz, 1998; Mark, Abrams, & Na, 2003), people must integrate the processing of the interruption itself and the resumption of the interrupted task. Thus, interruption enables adaptation, giving people opportunities to adjust their attention to a continuously changing environment. Yet, interruptions may also be disruptive, causing people difficulty in resuming an interrupted task or procedure. The investigation described in this paper recognizes that interruptions can be integral parts of continuing performance but may have negative consequences if not properly managed.

This is a unique investigation in the context of the psychological literature. Although past research suggests some fundamental principles for understanding interruptions, it ignores an entire category of tasks and behaviors. Only in recent years have researchers begun to ask questions about interruptions that are fast-paced, frequent, anticipated, and highly relevant to on-going overall performance. This investigation considers such interruptions and is based on a simple theoretical idea - people can learn to manage interruptions by strategically balancing the memory requirements associated with integrating the interruption with the interrupted task. A team decision making task, modeled after military aircraft classification tasks, serves as a dynamic and complex context in which to study the interruptions of interest. Here, messages among teammates are very important for accurate classification decisions. However, the nature of teamwork and a dynamic environment result in various levels of message timeliness and relevance, creating various types of interruption. A series of experiments was conducted to investigate the conditions under which people are able to effectively attend to interrupting messages and to quickly and accurately resume the interrupted decision task.

Interruption is pervasive in everyday life, for example, a car phone ringing, a colleague entering an office, an intelligent "office assistant" offering a suggestion, or instant messages arriving. One of the earliest studies of interruption suggested a relationship between interruption and people's task memory. Specifically, people remembered interrupted tasks better than non-interrupted tasks (Zeigarnik, 1927; Van Bergen, 1968). However, it is easy to understand at an emotional level that interruptions can be disruptive. Subsequent research on interruption and tasks switching indicates that the main consequence is longer time-on-task for the interrupted task (Gillie & Broadbent, 1989; Rogers & Monsell, 1995; Trafton, Altmann, Brock, & Mintz, 2003). This is often understood in terms of the task "resumption lag," the time added to the interrupted task as a result of needing to resume it following an interruption. Sometimes, people will even neglect the original task altogether (O'Conaill & Frohlich, 1995; Dismukes, et al., 1998) resulting in frustration, stress and a decreased sense of well-being (Kirmeyer, 1988; Zijlstra, Roe, Leonora, & Krediet, 1999; Adamczyk & Bailey, in press). Why are interruptions disruptive? In his review of the literature, McFarlane (1987) claims that interruptions are generally disruptive as the result of a person being distracted from the main task, forgetting what he or she was doing prior to the interruption (Adams & Pew, 1990; Preece, Rogers, Sharp, Benyon, Holland, & Carey, 1994).

Researchers have found a variety of characteristics that make interruptions most disruptive to task resumption. These include high interruption frequency (Spier, Valacich, & Vessey, 1999; Zijlstra, et al., 1999), similarity between interruptions and interrupted tasks, (Gillie & Broadbent, 1989; Cellier & Eyrolle, 1992; Edwards & Gronlund, 1998), task or interruption complexity (Gillie & Broadbent, 1989; Cellier & Eyrolle, 1992; McKnight & McKnight, 1993; Spier, et al., 1999; Zijlstra, et al., 1999), the relatedness of interruptions to the interrupted task (Spier, et al., 1999; Czerwinski, Cutrell, & Horvitz, 2000; Cutrell, Czerwinski, & Horvitz, 2000), and the type of activity occurring at the point of interruption (Czerwinski, et al., 2000; Cutrell, et al., 2000; Monk, Boehm-Davis, & Trafton, 2002). The experiments reported in the current research use a task that encompasses all of these characteristics. Interruptions are frequent and occur during a cognitively demanding decision task. In fact, the interruptions are so common that they are anticipated, and overall performance is dependent on the successful integration of interruptions into on-going task performance. There is a high degree of similarity among different decision events, and interruptions are highly related to the on-going decision task. Thus, task resumption is only a portion of performance. Memory for the interrupting task and integration of this memory with the resumed task is crucial to performance. The interruption is not merely secondary but an essential element of the "main" task. People must balance processing of the interruption itself with the resumption of the interrupted task.

Some of the literature has investigated a variety of ways in which people deal with interruptions. Some studies suggest that different personality characteristics determine who will be a good or poor interruption manager (Jolly & Reardon, 1985) and that user selection may be one way to minimize the impact of interruptions (Franke, Daniels, & McFarlane, 2002). Other research focuses on cooperative strategies to determine when messages should occur (Roger, Bull, & Smith, 1988; Dismukes, et al., 1998; McFarlane, 1999; Hudson, Christensen, Kellogg, & Erickson, 2002). Finally, some research has focused on the interface that people use to perform a task and process interruptions (Field, 1987; Arroyo, Selker, & Stouffs, 2002; Cutrell, Czerwinski, & Horvitz, 2001; Patten, et al., 2003). The research considered in this dissertation investigates interruption processing at a finer level. Experiments were designed to evaluate how people process interruptions in the context of the unique memory requirements imposed by interruptions that are frequent and necessary for task performance. The experimental situation is one in which people must continually process existing information and integrate incoming data in order to be successful.

Given the complex interplay between the main task and the interruption, is it possible to overcome disruption caused by an interruption so that it does not interfere with the interrupted task? At a very basic level, poor task resumption is a consequence of memory failure or interference (Stroop, 1935). Thus, it seems logical that an effort to improve memory for the task resumption point would improve interruption management and decrease task disruption. However, how people use their memories to manage interruptions has been a point of controversy in the literature. Gillie and Broadbent (1989) concluded that neither the length of the interruption nor the ability to control the point of the interruption improved task resumption. Their study participants did not use opportunities in the task to improve the speed with which the main task is resumed, presumably because they were not able to quickly recall the task resumption point.

In contrast, Altmann and Trafton (2002) suggested this finding was due to a lack of experimental control over the interruption and the imposition of other unexamined strategies. In fact, several memory strategies for resuming an interrupted task have been proposed in the literature. These include speeding up the primary task (Latorella, 1996; Zijlstra, et al., 1999), using external memory aids (Cutrell, et al., 2000), and postponing the interruption (Zijlstra, et al., 1999). Altmann and Trafton (2002) propose an alternative strategy, based on the fact that there is a "window of opportunity" (p. 65) during an interruption in which people are able to rehearse the task resumption point (see Figure 1).

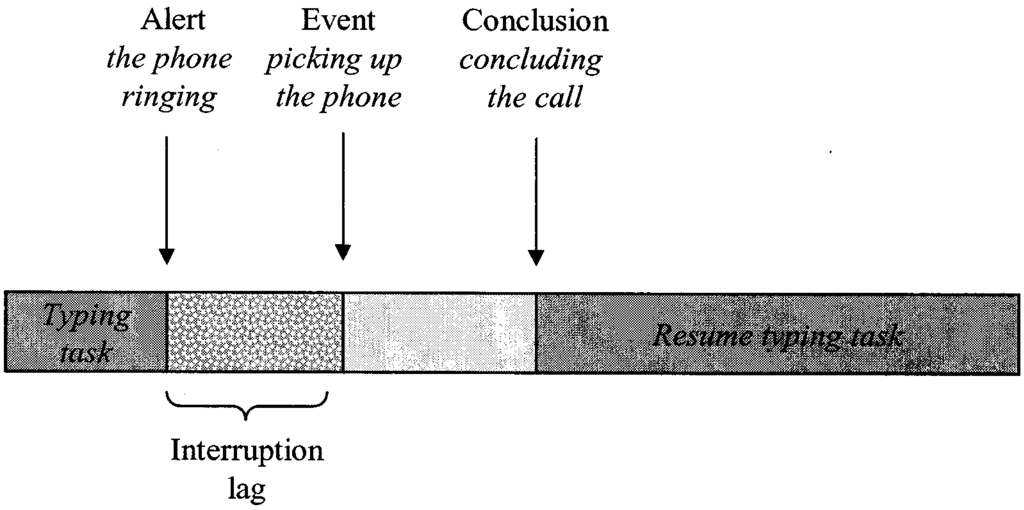


Figure 1. The anatomy of an interruption based on Altmann and Trafton (2002). For example, a phone call serves as an interruption to a typing task. The interruption lag, the period of time between the alert and picking up the phone, serves as a "window of opportunity" to rehearse the task resumption point.

In Altmann and Trafton's framework, an interruption has three main parts: an alert (e.g., the phone ringing), an event (e.g., picking up the phone), and a conclusion (e.g., hanging up the phone). A window of opportunity, or interruption lag (Altmann & Trafton, 2002; Trafton, et al., 2003) occurs between the interruption alert (e.g., the phone ringing) and the interruption event (e.g., picking up the phone). Although not explicitly indicated in Figure 1, Altmann and Trafton's framework also describes the resumption lag, the time added to the interrupted task as a result of needing to resume it following an interruption.

Interruptions in the form of messages are common but expected elements of teamwork. In a team communication task (Figure 2), this type of interruption is composed of an alert (e.g., receiving the message from a teammate), an event (e.g., reading the message), and a conclusion (e.g., dismissing the message). In this task, the interrupting event, the message must be attended to in order to assure optimal performance. Team decision-making is an area ripe with opportunities for studying interruptions. Team decision-making is a situation in which the interrupting task (communication from a teammate) is memory intensive. Relevant messages need to be remembered. Resuming the interrupted task (initial decision-making) is also memory intensive. Remembering the task resumption point is necessary for efficient task resumption. Anecdotal evidence indicates that team members using the "Team Argus" simulation (Schoelles & Gray, 2001) struggled with the requirement to share information in a timely and efficient way (Adelman, Yeo, & Miller, in preparation).

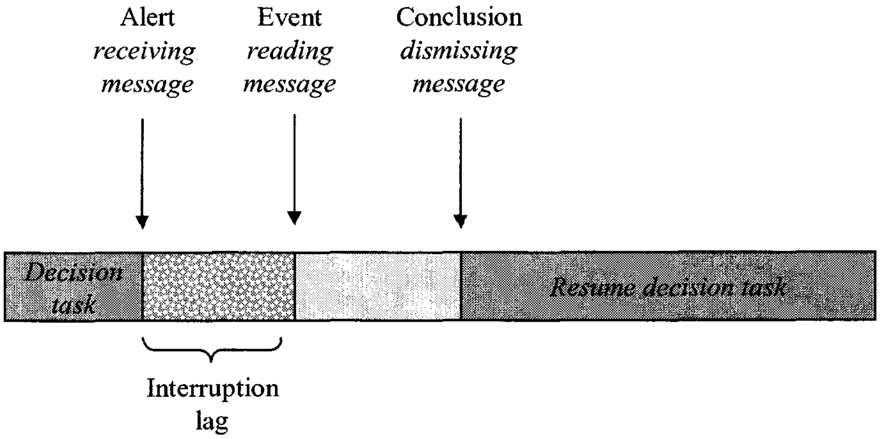
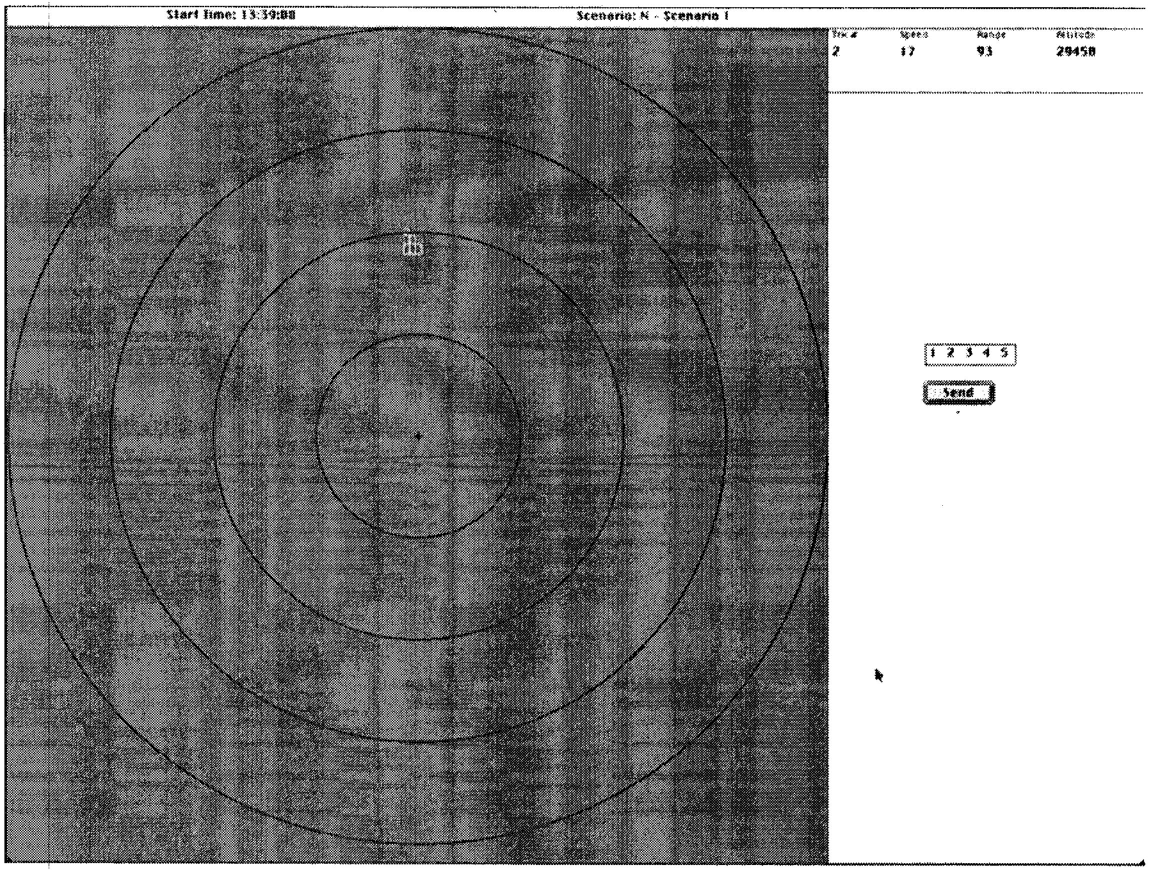


Figure 2. An example of an interruption in a team communication task. A message from a teammate serves as an interruption to a decision task. The interruption lag, the period of time between the alert and reading the message, serves as a "window of opportunity" to rehearse the task resumption point.

In this situation, people must process the interruption by balancing the memory requirements of the interrupting message and the resumption of the interrupted task. An experiment was designed to evaluate this challenge. The experimental situation used in this research, called "Team Argus," is one in which people must continually process existing information and integrate incoming data in order to be successful. The first part of Team Argus requires that interrupting messages be remembered after they were dismissed. This message information would either be needed immediately or for a future aircraft decision Thus the experimental design allows the evaluation of interruption performance over time (on subsequent tasks). The second part of Team Argus requires that the interrupted task be resumed in a timely and accurate manner. Based on Altmann and Trafton's (2002) predictions, decision-makers who use the interruption lag to rehearse where they leave off in the interrupted task will be better able to resume performance of the interrupted task compared to those who use no such strategy (in other words, reducing the task resumption lag). To test this, instruction to rehearse the task resumption point was experimentally manipulated. Rehearsing the task resumption point is a modest solution to the problem of interruptions, but was expected to have large consequences in alleviating task disruption. Thus the experimental design also allows the evaluation of current task resumption performance. Conclusions therefore can be drawn about the way that people balance the memory requirements of the interrupting message and the resumption of the interrupted task.

2. I-ARGUS TASK DESCRIPTION

I-Argus, or "Interruptible Argus" was developed to build on work conducted using "Team Argus" (Schoelles & Gray, 2001) and to create an environment in which to study interruptions and task resumption. I-Argus is a simulation of a team radar operator decision-making task (see Figure 3). It is shown here in black and white, but study participants saw it in color. An operator works individually at a computer to assess each aircraft that appears on the radar scope and to make recommendations to a "leader" about what course of action to take. Although the team was simulated (i.e., operators work with experimentally-generated computerized teammates), the behavior of the simulated teammates mirrored that of real teams using "Team Argus" (Adelman, Miller, Henderson, & Schoelles, 2003; Adelman, Miller, & Yeo, 2004).



Trk #	Speed	Range	Altitude
2	17	93	29450

Figure 3. I-Argus as seen by a team member. In this example, an icon on the left-hand radar scope represents aircraft #2. The aircraft has been hooked by the team member revealing aircraft #2 information in the upper right-hand corner of the screen. The team member will select a decision (a course of action 1-5) on the right-hand side of the screen.

I-Argus aircraft are displayed on a geocentric radar display. Operators see one aircraft at a time on the radar scope and issue a decision for each one. Figure 4 depicts the serial nature of the task.

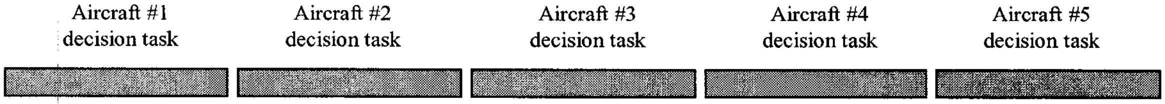
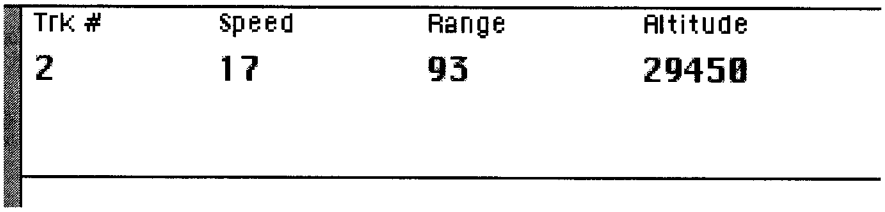


Figure 4. Decisions are made one at a time for each aircraft.

When an aircraft appears on the screen, an operator "hooks" it by clicking on it to display three pieces of aircraft data (range, speed, and altitude) in the upper right-hand corner of the display (see Figure 5). Each piece of data was assessed by the participant in terms of its threat level on a 0-3 scale (see Appendix B, e.g., speed of 600 knots is highly threatening or 3). Overall threat for the aircraft was determined by adding together the threat level of all the data (see Appendix C). This produces an overall numeric score that corresponds to a course of action.



Trk #	Speed	Range	Altitude
2	17	93	29450

Figure 5. Three pieces of data (range, speed, and altitude) appear when the team member hooks aircraft #2.

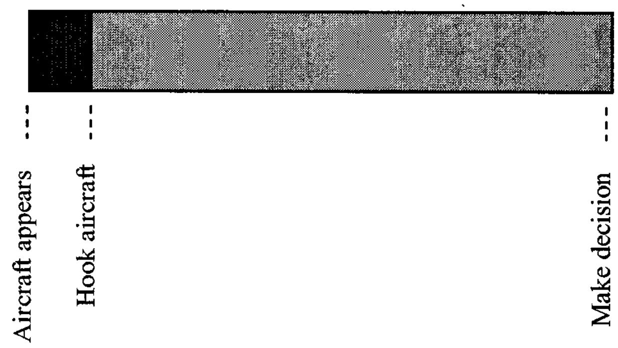


Figure 6. Each aircraft decision is composed of two subtasks, hooking the aircraft icon after it appears and making a decision.

Figure 6 depicts the task process. The decision-making component involves reading data values for speed, range, and altitude and performing calculations to produce an overall score. Details regarding these calculations may be found in Appendix D. After calculating a decision, the operator sends it to the team "leader" by selecting a number between one and five and clicking the send button in the lower right-hand corner of the screen (Figure 7).

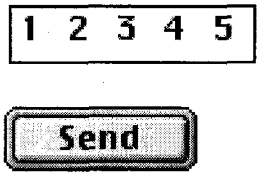
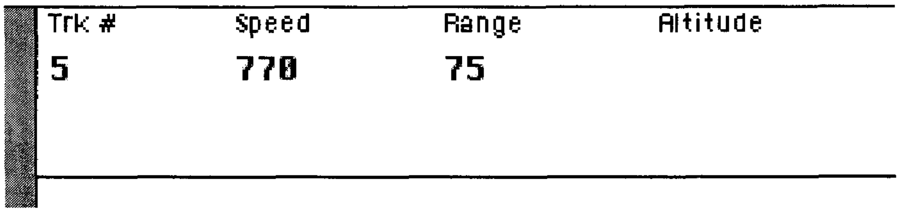


Figure 7. The team member issues a decision by selecting a number between one and five that equates to a course of action and clicking the send button.



Trk #	Speed	Range	Altitude
5	770	75	

Figure 8. Occasionally, aircraft information is missing (e.g., aircraft #5 altitude).

Occasionally, a piece of data (e.g., altitude) is unavailable for a particular aircraft (Figure 8). This mimics the real world situation of inconsistent or unreliable data, and provides an experimental opportunity for communication between teammates in the form of messages from a teammate. Participants were instructed to make the best decision that they could with missing data. However, sometimes the missing piece of data is received in the form of a message from a teammate (i.e., an experimentally-generated message). These messages served as interruptions during decision making; the messages contained a piece of data (e.g., altitude) that was relevant to some aircraft. Hence, the terms "messages" and "interruptions" can be used interchangeably in the context of this paper.

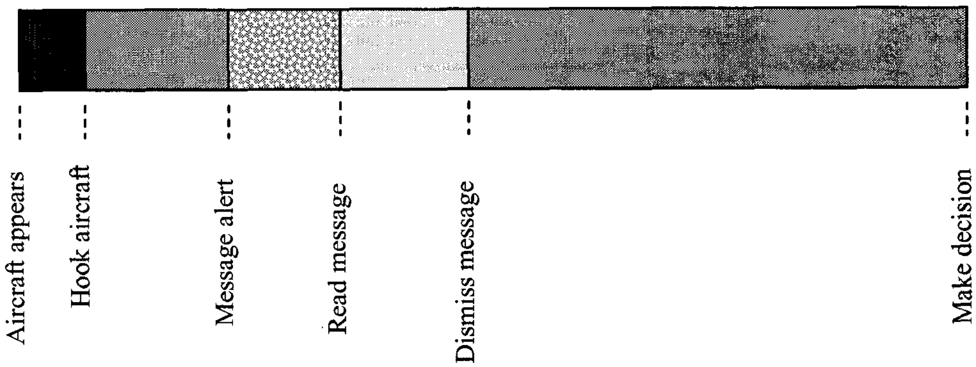
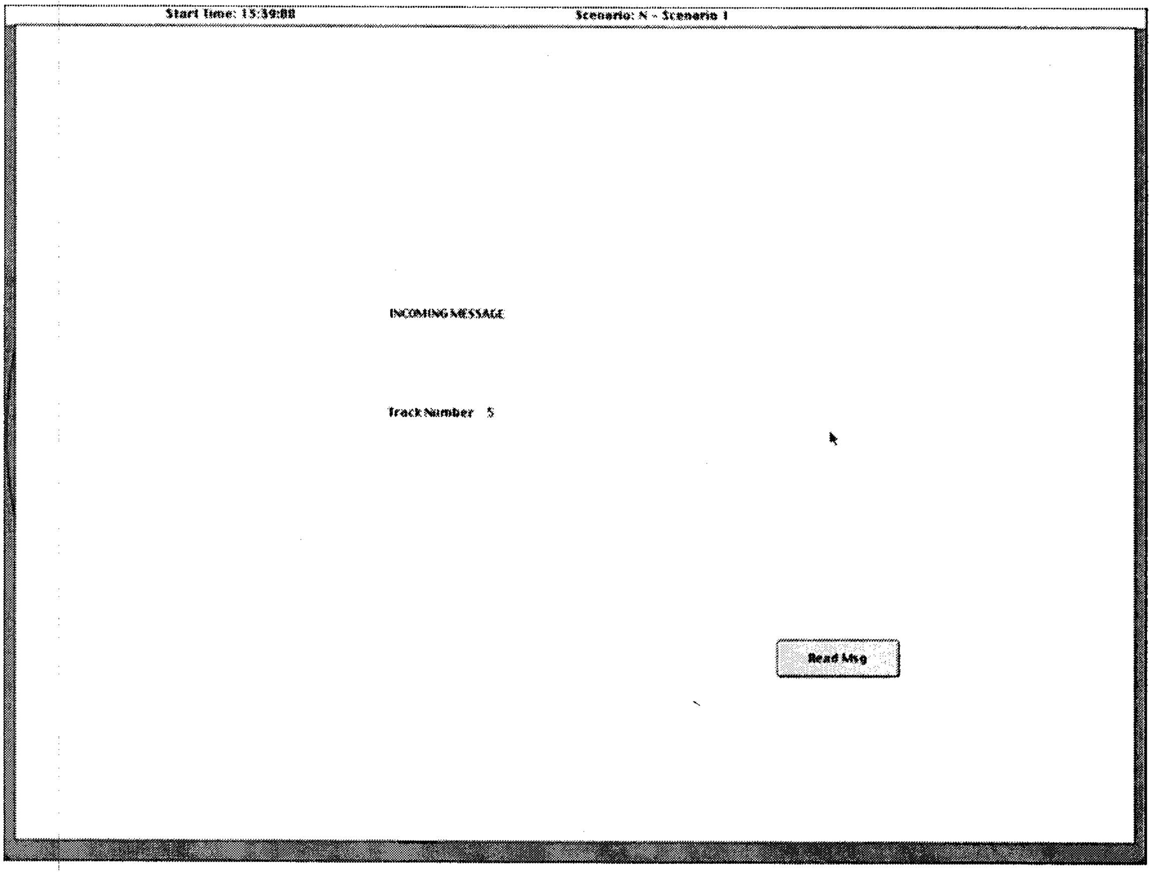


Figure 9. Occasionally, an aircraft decision is interrupted by a message.

Altmann and Trafton (2002) provide a framework to describe interruptions to aircraft decision making (Figure 2). Each interruption is composed of an interruption lag and the reading of the message itself (Figure 9). The interruption lag begins when an interruption alert hides the radar screen from view and continues until the operator decides to attend to the interruption and reads the message content by clicking on the read button (see Figure 10a). Half of the experimental participants were explicitly instructed to use the interruption lag to rehearse the task resumption point. Half received no instruction regarding the interruption lag. After selecting to read the message, the message content appears giving the operator a piece of data about some aircraft (see Figure 10b). When the operator dismisses the message, the interrupted aircraft reappears on the screen and the message is no longer available. Therefore, operators need to remember the message data while they resume the interrupted decision making and while making decisions on subsequent aircraft.

a)



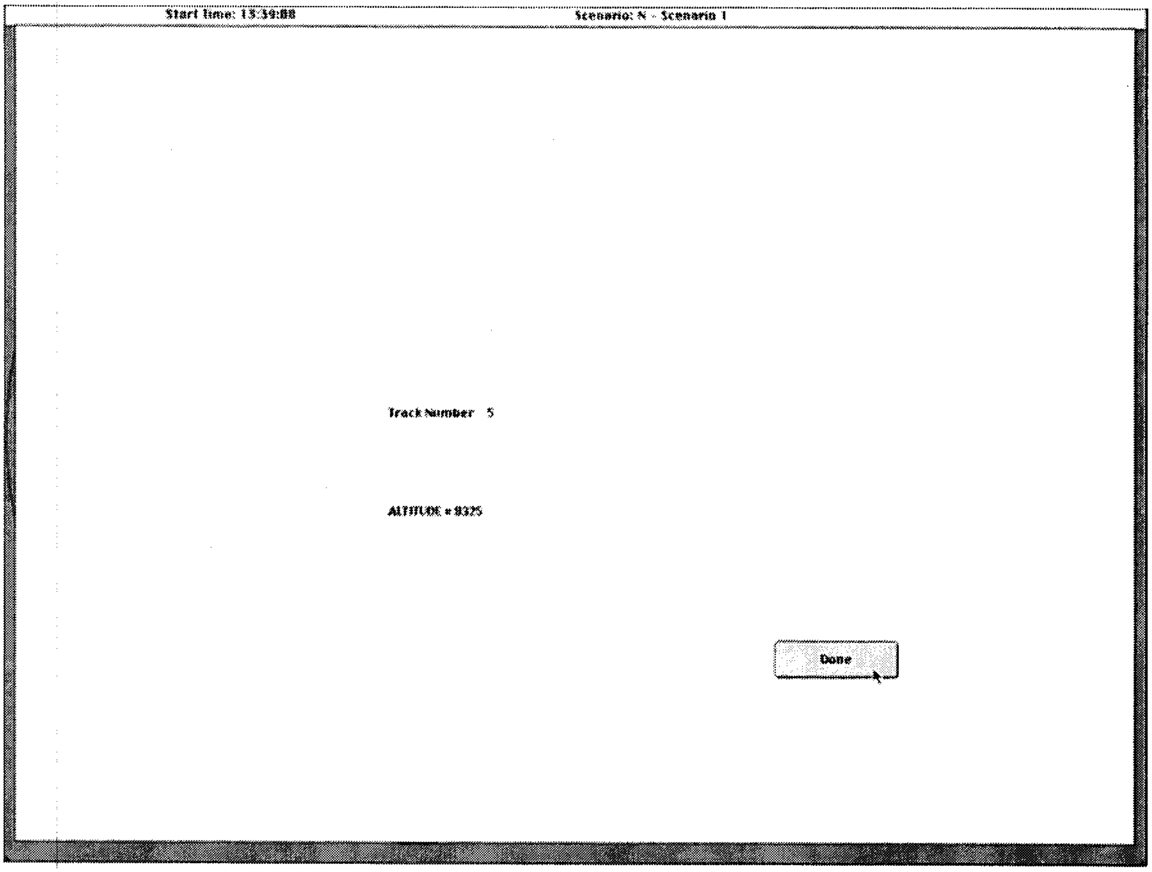
Start Time: 15:39:08

Scenario: N - Scenario 1

INCOMING MESSAGE

Track Number 5

b)



Start time: 13:39:08

Scenario: N - Scenario 1

Track Number 5

ALTITUDE = 8325

Figure 10. a) The screen seen during the interruption lag. b) The screen following the interruption lag contains the message content (aircraft identifier and data, e.g., altitude).

Interruptions are messages that contain information (e.g., data such as altitude) that must be remembered in order to make accurate decisions. Sometimes these messages will contain information relevant to the current aircraft for which a decision is being made (in this case, aircraft #5 as seen in Figure 11. In this case, the information contained in the message must be remembered only for as long as it takes to dismiss the message and resume the interrupted task.

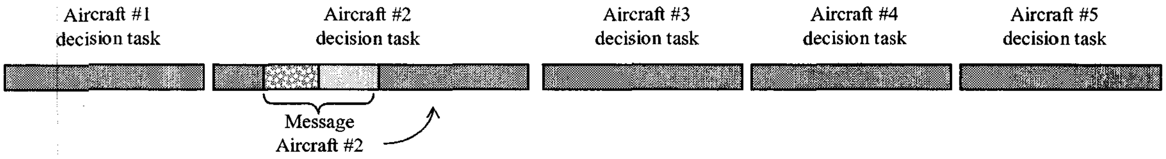


Figure 11. Occasionally, an aircraft decision is interrupted by a message. This message may be immediately relevant for the interrupted aircraft decision. These will be referred to as Interrupted-Utilize aircraft.

However, sometimes the interrupting message is not immediately relevant. Instead, it contains information (e.g., data such as speed or altitude) that is relevant to some aircraft that has not yet appeared on the screen as seen in Figure 12. In this case, the message must be remembered during resumption of the interrupted task but also during any subsequent aircraft (these will be referred to as Memory Load aircraft). When the relevant aircraft appears, the remembered information can then be used to make a decision (these will be referred to as Recall Message aircraft).

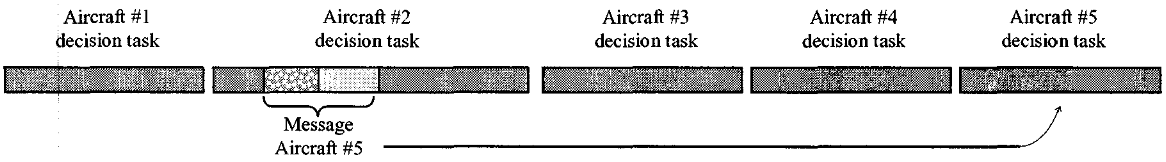


Figure 12. Occasionally, an aircraft decision is interrupted by a message. This message may be relevant for some future aircraft decision. These will be referred to as Interrupted-Learn aircraft

Structuring the task in the ways outlined above enables experimental evaluation of six categories of aircraft, defined by the a) amount of information initially available for that aircraft, b) whether or not decision making for that aircraft was interrupted, and c) whether the information contained in the interrupting message is immediately relevant or relevant in the future (Table 1).

Table 1. Six types of aircraft appear in I-Argus.

	Experimental Aircraft		Control Aircraft
Information	**Interruption**	**No Interruption**	**No Interruption**
Complete	Interrupted-Learn (Read message for future use).	Memory Load (Remember message for future use).	Complete (Baseline best decision).
Incomplete	Interrupted-Utilize (Read message for immediate use).	Recall Message (Use message read earlier).	Incomplete (Baseline worst decision).

Two types of aircraft (Complete and Incomplete) served as baseline performance tasks. Evaluation of these aircraft addresses the following question: How well do people make decisions when they have complete data and when they have missing data? An aircraft is complete when range, speed, and altitude are all available to the decision maker. An aircraft is incomplete when one piece of data is not available. The next type of aircraft (Interrupted–Learn) was used to investigate the impact of interruptions on decision performance. These aircraft are presented with all three pieces of data; the interruption message contains information relevant to a future aircraft. In the Memory Load case, the aircraft is presented with all three pieces of data and decision-making is not interrupted, but the operator is holding information in memory for a future aircraft. In the Recall Message case, when the relevant aircraft appears on the screen, it has incomplete data and the operator must remember and use the earlier message content in order to have all three pieces of information; this decision-making process is not interrupted. A final type of aircraft (Interrupted-Utilize) was used to assess cases when the message content is immediately relevant to the aircraft task; that is, the message presented during the interruption contains information relevant to the current aircraft.

The variety of aircraft represented in I-ARGUS enables the exploration of how people process interruptions in realistic, complex situations. Here, interruptions are necessary and frequent aspects of on-going performance. Thus, interruptions cannot be ignored. However, they must be processed to minimize disruption to the interrupted task. The interrupting task and task resumption are both memory-intensive. How do decision-makers balance the interrupting task (remembering a message) with task resumption (remembering the task resumption point)?

3. EXPERIMENT 1

The first experiment was designed to test how decision-makers process interruptions that are necessary and frequent. First, how do interruptions affect performance in a complex task? The structure of the experimental task is unique in that it allows the observation of disruption to both interrupted decisions and subsequent decisions including those for which the interruption is relevant. Second, how do people balance attending to and remembering the interruption with remembering how to resume the interrupted task? We believe that people will be challenged by the requirement to balance the interruption (remembering the message) with the task resumption (remembering the task resumption point). The experimental situation was designed to interleave interrupting messages with a series of aircraft decision tasks. In analyzing performance, the following measures will be considered.

Decision time. This is the time spent making a decision. It is measured as the time between hooking the aircraft and clicking the "Send" button to complete a decision. For interrupted decisions, the interruption lag and message time are subtracted out of the total time. Thus, interrupted decision time was composed of two parts, the time prior to the interruption and the time after the interruption. Decision time measures baseline decision-making on non-interrupted aircraft (Complete and Incomplete). This will be compared to aircraft that may be disrupted by an interruption (Interrupted, Memory Load, and Recall Message). It is expected that interruptions will increase total decision time when compared to baseline performance. Note that the task resumption lag is not explicitly measured in this task. However, the resumption lag is assumed to be part of overall time-on-task and is reflected by the differences between interrupted and non-interrupted aircraft time-on-task.

Accuracy. This is the correctness of participants' decisions, or the proportion of correct decisions out of total decisions. It is expected that accuracy may suffer compared to baseline performance as a result of forgetting an interrupting message that is needed for an aircraft decision (Recall Message).

Interruption lag. For decisions that are interrupted, this is the time spent switching attention from the initial decision task to the interrupting message content (Altmann and Trafton, 2002). It is measured as the time between the message alert (the message is received) and reading the message (clicking on the "Read" button). The interruption lag will be used to measure the amount of time decision-makers spend switching attention to the interruptions and thus may reflect time rehearsing the task resumption point. It is expected that interruption lags will be longer for participants instructed to rehearse the task resumption point and that this will be associated with shorter decision times for Interrupted aircraft indicating less disruption from the interruption.

Message time. This is the time spent reading a message. It is measured as the time between choosing to read the message (clicking the "Read" button) and dismissing the message (clicking the "Done" button).

METHOD

Participants

Twenty-four undergraduates volunteered to participate for two hours in exchange for course credit. Twelve men and twelve women participated, and their median age was 19.98 years (SD=2.04). Participants were randomly assigned to one of two rehearsal strategy conditions - instructed or not instructed. Half were instructed to rehearse the task resumption point during the interruption lag and half were not. Participants were also assigned to receive one of two message types, speed or altitude. Four participants were randomly assigned to serve as protocol participants. These participants were videotaped while talking aloud during the task.

Procedure

At the beginning of each session, the participant was greeted and seated at a Macintosh G3 computer. He or she then provided informed consent and answered a brief set of demographic questions.

Pre-test

Because individual memory differences might influence the use of the rehearsal strategy, the n-back task was administered as a pretest to assess working memory capacity (McElree, 2001). Numbers 0-9 were presented on the computer screen one at a time (20 numbers in each trial). Participants were asked to report whether the number on the screen was the same or different as one of the previous numbers - the last number, 2 numbers back, or 3 numbers back. After reaching criteria of 95% on the 1-back practice task, each participant performed 3 trials each of the 2-back and 3-back task. The proportion correct was averaged across all trials (2-back and 3-back) to produce an overall score.

Training

Participants received verbal and written instructions for performing the experimental task. They were also encouraged to ask questions of the researcher during the task. First, they spent approximately 20 minutes reading the instructions, which were supplemented with self quizzes. The instructions described the task, the team environment, and how to accomplish the task (see Appendix A).

In order to assure that participants would not ignore messages, participants were instructed to be as accurate as possible in their decision-making. They were told that the length of their scenarios would depend on their accuracy; less accurate performance would result in a greater number of aircraft. It was expected that participants would be motivated to complete the study quickly and thus they would try to pay attention to all messages and make accurate decisions. However, accuracy did not actually affect the number of trials participants performed.

Following written and verbal instructions, participants were asked to perform practice trials in sets of 10 aircraft decisions each. Participants first completed a scenario containing 10 aircraft and no messages. Participants who made fewer than 7 accurate decisions repeated an isomorphic scenario of 10 aircraft. When each participant reached criterion, they were given a second 10 aircraft scenario containing 3 interruptions, messages containing a piece of data associated with some future aircraft. When each participant reached criterion (7/10 correct), he or she was allowed to move on to the experimental trials. The four protocol participants also received brief training in providing a verbal protocol. The process of "talking aloud" during the experimental task was described to them, and they practiced with the experimenter by performing a series of paper and pencil addition and subtraction problems. Each protocol participant was asked to talk aloud while performing the task and was told that the experimenter would remind them to "keep talking" occasionally. Each protocol was videotaped.

Experimental Scenarios

Following a break, each participant performed two isomorphic scenarios each containing 64 aircraft on the I-Argus system. Aircraft arrived 4 seconds after a previous aircraft decision was issued. Scenarios were counterbalanced between participants to control for order effects.

Aircraft interruptions differed in terms of timing and relevance. Interruption timing was defined as the time (in seconds) after the aircraft appeared that the message would arrive. Interruptions occurred 1, 3, 5, or 7 seconds after the aircraft appeared. Timing varied randomly within a scenario, but each timing level occurred 4 times. Interruption relevance was defined as number of aircraft into the future for which a message is relevant. Each aircraft interruption was relevant to 0, 1, 2, or 3 aircraft ahead. If relevance was equal to 0, then the message was immediately relevant, if it was equal to 1, then the message was relevant to the next aircraft, and so on. Relevance varied randomly within a scenario. Table 2 provides an example of a series of aircraft in a scenario. Because of the small number of instances for each of these different types of interruption, the differences will not be considered in the forthcoming analyses.

Table 2. A scenario example; each aircraft arrives 4 seconds after a previous decision. Here, a message interrupts aircraft #18. The message arrives 5 seconds after the aircraft appears and is relevant to two aircraft ahead, aircraft #20.

Aircraft#	**Time (sec)**	**Range**	**Speed**	**Altitude**	**(Message)**	**Aircraft Type**
...						
17	+4	23	465	20000		Complete
18	+4	47	220	16003		Interrupted-Learn
	+5			27010	(Aircraft#20)	
19	+4	82	550	3000		Memory Load
20	+4	130	159		←	Recall Message
21	+4	66	230	4500		Complete
22	+4	48	780			Incomplete

Debriefing

After the second experimental scenario, each participant was asked a series of questions about the task and then was debriefed.

RESULTS AND DISCUSSION

Experiment 1 was designed to test the effect of interleaving interruptions with on-going decision-making. Thus, analyses compared performance on the six types of aircraft detailed in Table 1 (Complete, Incomplete, Interrupted-Learn, Memory Load, Recall Message, and Interrupted-Utilize). Message type (speed or altitude data) and working memory pre-test scores were not significantly related to performance and were collapsed in subsequent analyses. In all cases, reaction time data were trimmed to account for outliers. Standard errors of means were calculated. Individual reaction times greater or less than four times the standard error of the mean were trimmed from the sample. Figure 13 provides an overview of differences in decision-making processes for interrupted and non-interrupted aircraft in Experiment 1. Time-on-task is shown for each step in an aircraft decision including hooking (or clicking) an aircraft icon after it appears in order to see its data and deciding on a course of action. When an aircraft is interrupted, interruption lag time and message time are also represented.

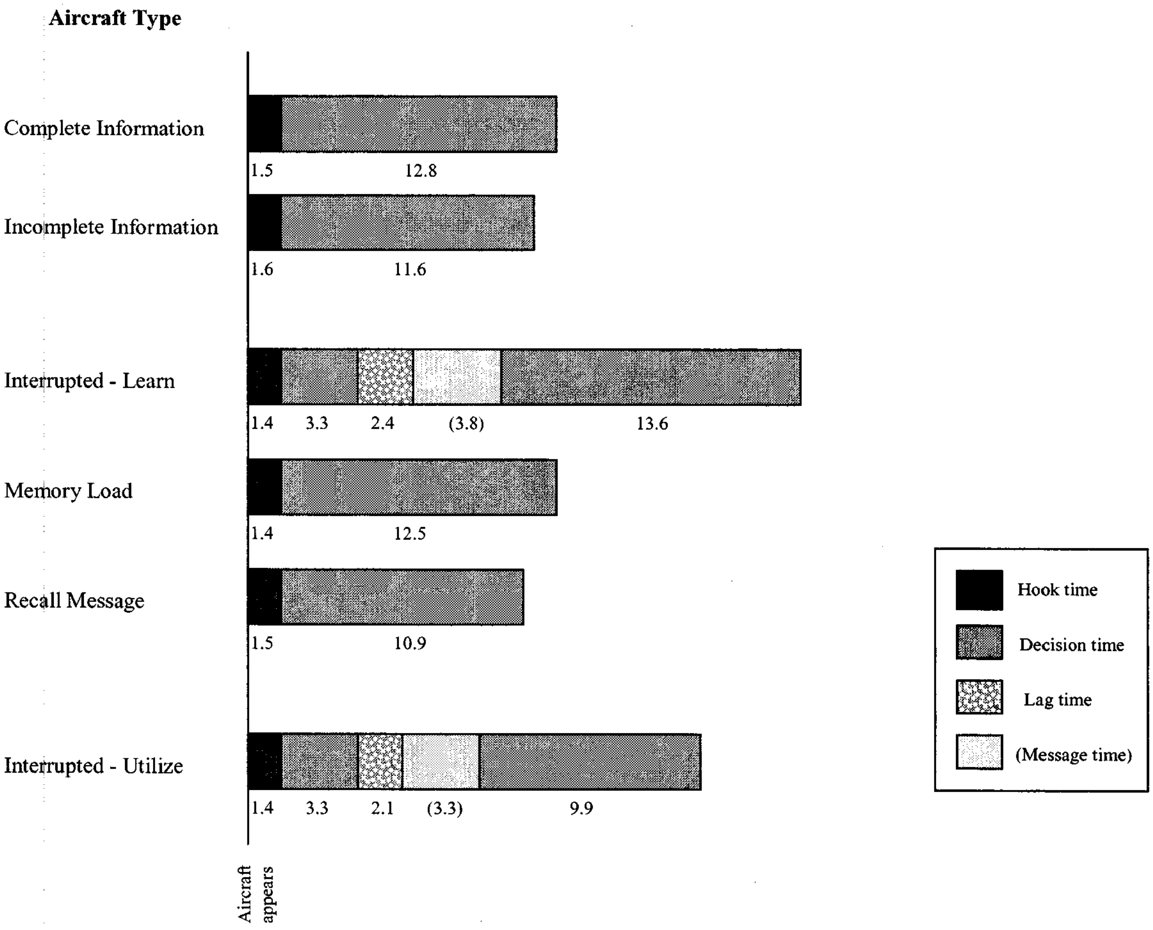


Figure 13. Time-on-task for different aircraft types in Experiment 1.

The I-ARGUS decision task was designed to evaluate two aspects of interruption processing ignored in the literature. First, what is the impact of frequent and necessary interruptions on decision-making, both immediately and during subsequent decisions? Second, how do people balance the memory requirements imposed by the interruption itself (remembering the message) and task resumption (remembering the task resumption point)? These questions can be answered by comparing performance across the six different types of aircraft represented in Figure 13.

What is the impact of frequent and necessary interruptions on decision-making, both immediately and during subsequent decisions? Differences among aircraft types were present for both accuracy (proportion correct) and decision time (in seconds). Interruptions' effects on decision time are the primary measure of interest in this experiment. However, where appropriate, decision accuracy is used to understand these effects. The instructed rehearsal strategy did not affect accuracy, $F(1,22)=0.452$, $p=0.508$ and only marginally affected decision time, $F(1,22)=3.120$, $p=0.091$. There was no interaction between rehearsal strategy and aircraft type for accuracy, $F(5,110)=1.016$, $p=0.391$ or decision time, $F(5,110)=1.908$, $p=0.112$. However, type of aircraft did affect accuracy, $F(5,110)=13.862$, $p<0.001$, and decision time, $F(5,110)=33.186$, $p<0.001$. Type of aircraft affected accuracy (Figure 14) and decision time (Figure 15). Tukey's Honestly Significant Difference (HSD) tests were used to evaluate differences among treatment means at $p=0.05$. These differences are shown in Tables 3 and 4.

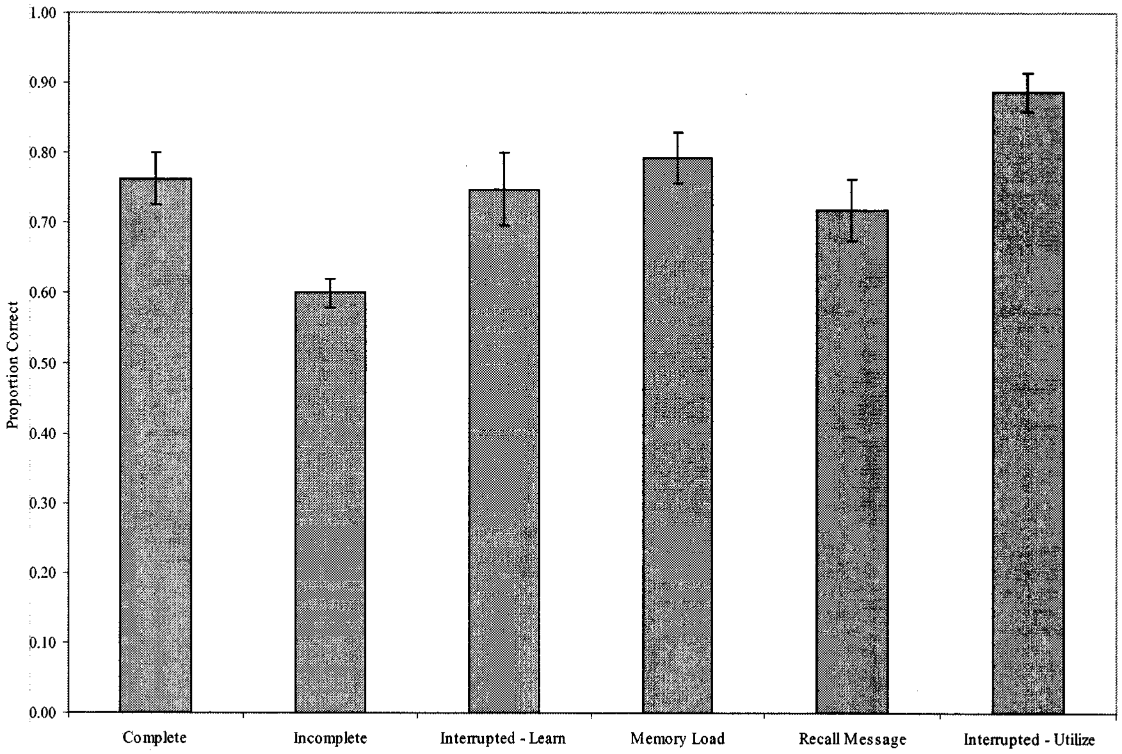


Figure 14. Accuracy for aircraft decisions in Experiment 1.

Table 3. Differences among Experiment 1 accuracy scores (proportion correct).

		Complete	Incomplete	Interrupted-Learn	Memory Load	Recall Message	Interrupted-Utilize
		0.76	0.60	0.75	0.79	0.72	0.89
Complete	0.76		0.16*	0.02	0.03	0.04	0.12
Incomplete	0.60			0.15*	0.19*	0.12	0.29*
Interrupted-Learn	0.75				0.05	0.03	0.14*
Memory Load	0.79					0.07	0.10
Recall Message	0.72						0.17*
Interrupted-Utilize	0.89						

Tukey's Honestly Significant Difference (HSD) = 0.14

* p=0.05

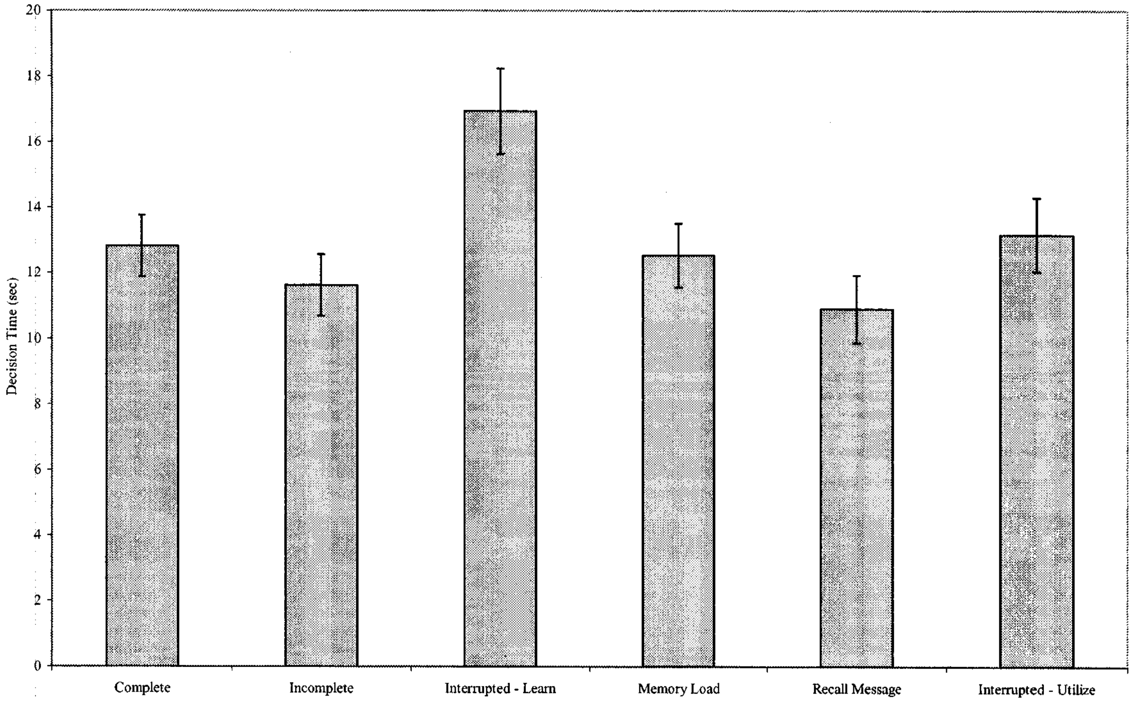


Figure 15. Decision time for aircraft in Experiment 1.

Table 4. Differences among Experiment 1 decision times (seconds).

		Complete	Incomplete	Interrupted-Learn	Memory Load	Recall Message	Interrupted-Utilize
		12.82	11.63	16.95	12.54	10.90	13.17
Complete	12.82		1.19	4.13*	0.28	1.92*	0.35
Incomplete	11.63			5.33*	0.91	0.72	1.55*
Interrupted-Learn	16.95				4.41*	6.05*	3.78*
Memory Load	12.54					1.64*	0.63
Recall Message	10.90						2.27*
Interrupted-Utilize	13.17						

Tukey's Honestly Significant Difference (HSD) = 1.50

* p=0.05

Baseline performance can be assessed by looking at Complete and Incomplete aircraft. Complete decisions (M=12.82 seconds, SD=3.40) took longer than Incomplete decisions (M=11.63, SD=3.36), but this difference is not significant, HSD=1.5. However, average accuracy for Complete aircraft decisions was 0.76 (SD=0.18), significantly higher than 0.60 (SD=0.10) for Incomplete aircraft decisions, HSD=0.135. Thus, missing a single piece of decision data resulted in less accurate decisions.

These two types of aircraft, Complete and Incomplete, serve as anchors for evaluating decision performance for all other aircraft, so that the impact of interruptions on decision-making can be assessed. Two types of interrupting messages were present, those messages which were immediately relevant to the interrupted aircraft and those that were relevant to some future aircraft. Interrupted-Utilize (immediately relevant) decision times (M=13.17, SD=4.35) were statistically the same as Complete aircraft (12.82), HSD=1.5. Interrupted-Learn decisions (M=16.95, SD=4.81) took longer than both of these by approximately 4 seconds), HSD=1.5. Therefore, an interruption only seems to disrupt decision time when it requires considerable effort – remembering a message for subsequent use. Interestingly, Interrupted-Utilize decisions appear more accurate (M=0.89, SD=0.03) than Complete decisions (0.76) but this is not a significant difference, HSD=0.135. However, Interrupted-Utilize decisions were more accurate than Interrupted-Learn decisions (M=0.75, SD=0.05). Therefore, interruptions immediately relevant to the task at hand may have actually provided a boost to decision accuracy.

Interrupted decisions for which the message was immediately relevant were faster and more accurate than interruptions for which the message had future relevance. Surprisingly, these Interrupted-Utilize decisions were also no less accurate nor slower than baseline Complete aircraft. This finding is a unique contribution because the effect of integrating interruptions with the task has not been studied before. Interrupted-Learn aircraft required participants to remember the interrupting message for future use. This resulted in disruption to decision time. How did these interruptions affect subsequent decision making? Recall Message aircraft are those for which the earlier message was relevant; and Memory Load aircraft are those that arrive between the Interrupted aircraft and Recall Message aircraft.

Accuracy for Recall Message aircraft was 0.72 (SD=0.21); this was statistically no different than either Incomplete (0.60) or Complete (0.76) aircraft, HSD=0.135. These Recall Message decisions (M=10.9, SD=3.58) took less time than Complete decisions (12.82), HSD=1.5. Thus, remembering a previously presented piece of information was quicker than having all the information presented at once as in Complete aircraft. This suggests that people might have been able to speed up their subsequent use of the interrupting message information. In fact, follow-up interviews and a review of participant protocols suggest that people assessed the threat of message information at the time that the message was received, rather than when it became relevant. Thus, in processing a message people would remember the threat rating 0, 1, 2, or 3. Participants' protocols and reports indicated that this was easier than remembering raw data values (e.g., remembering an altitude threat rating of 0 is easier than remembering a raw altitude value of 25,400). This is an interesting adaptation to the demands of the task. Remembering a message over time appears to have been challenging enough to decision-makers that they developed a shortcut to simplify the memory requirements of the task.

Evaluations of message time support this interpretation. The time spent reading message content varied across aircraft type, F(1,22)=4.997, p=0.036. It did not vary across rehearsal strategy, F(1,22)=0.579, p=0.456, nor was there an interaction, F(1,22)=0.530, p=0.474. Participants spent less time reading messages when they were immediately relevant (M=3.31, SD=1.96) than when messages were relevant for some future aircraft (M=3.84, SD=1.88). Thus, when participants recognized that messages would not be immediately relevant to the interrupted task, they spent extra time processing the message. As noted above, it appears that this extra time was used to assign a threat rating to the raw data value, as this would be easier to remember. These results suggest that participants paid attention to messages and that they developed strategies to remember and use the message content when needed.

Memory load aircraft (M=0.79, SD=0.17), during which participants were holding this message content in memory, were statistically as accurate as Complete decisions (0.76) and more accurate than incomplete decisions (0.60), HSD=0.135. Additionally, Complete (12.82) and Memory Load decisions (M=12.54, SD=3.57) took the same amount of time, HSD=1.5. Thus, decision time and accuracy seems unaffected by holding a piece of information in memory. It also implies that participants did not spend time rehearsing the information during a decision, and this is supported by the behavior of the protocol participants. Participants instead may have used opportunities between each aircraft decision to rehearse this information. However, protocols do not offer evidence of this.

Thus far, the analyses indicate that some interruptions are disruptive to decision-making. Interrupting messages that must be remembered over time result in longest decision times on the interrupted task. Remembering interrupting messages did not affect subsequent performance on other aircraft decisions until the remembered data was needed. While people lost time on the initial interrupted task, it appears that they are able to save time later when that remembered data was used. Previous research has not examined the memory trade-offs associated with interruption, making this finding a unique contribution to the literature.

How else did people manage the memory requirements associated with interrupted decisions? It was expected that rehearsing the task resumption point prior to an interruption would alleviate the impact of an interruption on decision time. Half the participants were instructed to rehearse during the interruption lag, while half received no such instruction. As reported above, no significant interaction existed between aircraft type and strategy instruction condition (decision time, F(5,110)=1.908, p=0.112). However, because a difference was expected, interruptions were examined more closely. Interrupted-Learn decision times differed in terms of strategy condition, t(138)=2.121, p=0.036, as did Interrupted-Utilize decisions, t(138)=2.547, p=0.018. For both types of interruptions, participants instructed in the rehearsal strategy performed differently when interrupted than uninstructed participants. However, examining decision time means (see Figure 16) shows that this effect was opposite the direction predicted. For both types of interrupted aircraft, decisions were longer by participants who were instructed in the rehearsal strategy.

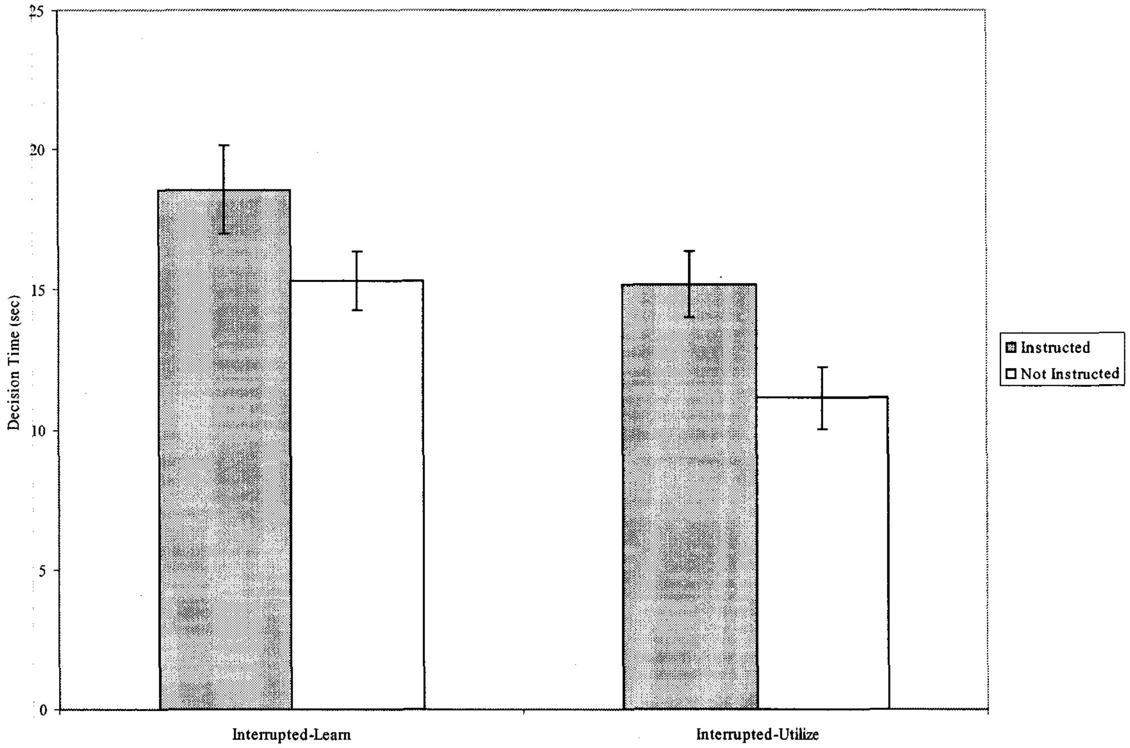


Figure 16. Average decision time was longer for participants instructed to use the rehearsal strategy for both types of interruption.

It took instructed participants approximately 4 seconds longer to complete interrupted decisions than participants who were not instructed. Therefore, the strategy to rehearse the task resumption point was unsuccessful in reducing task disruption and may have actually made decision times longer. This was especially true when interrupting message content needed to be remembered well into the future. Why? One possible explanation is that people did not use the strategy as they were instructed. Perhaps, instructed participants did something unexpected to increase their interrupted decision times. In order to investigate rehearsal strategy use, interruption lag time was examined.

For decisions that are interrupted, lag time is the time spent switching attention from the initial decision task to the interrupting message content. Instructed participants should have used this time to encode the task resumption point using the rehearsal strategy. Thus, lag times for instructed participants should be longer than for participants who were not instructed in this strategy.

There was a main effect of aircraft type, $F(1,22)=18.633$, $p<0.001$. Lag times were longer for Interrupted-Learn aircraft ($M=2.44$, $SD=0.846$) than for Interrupted-Utilize aircraft ($M=2.14$, $SD=0.625$). This finding is consistent with the notion that interruptions that must be remembered for future use are more disruptive than interruptions that are immediately relevant. Participants did not use as much time to prepare for task resumption when the message was immediately relevant as when the message was relevant to a subsequent aircraft. Lag times were longer for instructed participants ($M=2.52$, $SD=0.75$) than for non-instructed participants ($M=2.05$, $SD=0.69$). This difference is not significant, $F(1,22)=2.911$, $p=0.102$, but it is a trend consistent with the expectation that rehearsal strategy instruction led to an increase in the interruption lag. This indicates that instructed participants might have used the interruption lag to do some extra processing of the interrupted task when the message was not going to be immediately helpful.

However, there also was an interaction of aircraft type and rehearsal strategy instruction for lag time, $F(1,22)=13.019$, $p=0.002$ (see Figure 17). Lag times were significantly longer for instructed participants ($M=2.80$, $SD=0.82$) than for non-instructed participants ($M=2.07$, $SD=0.73$) for Interrupted-Learn aircraft only, $HSD=0.27$. This is likely related to the fact that for Interrupted-Utilize aircraft, interruptions were generally not disruptive. However, when messages were not immediately relevant, instructed participants had longer interruption lags and thus were doing some extra preparation related to diverting attention from the interrupted task.

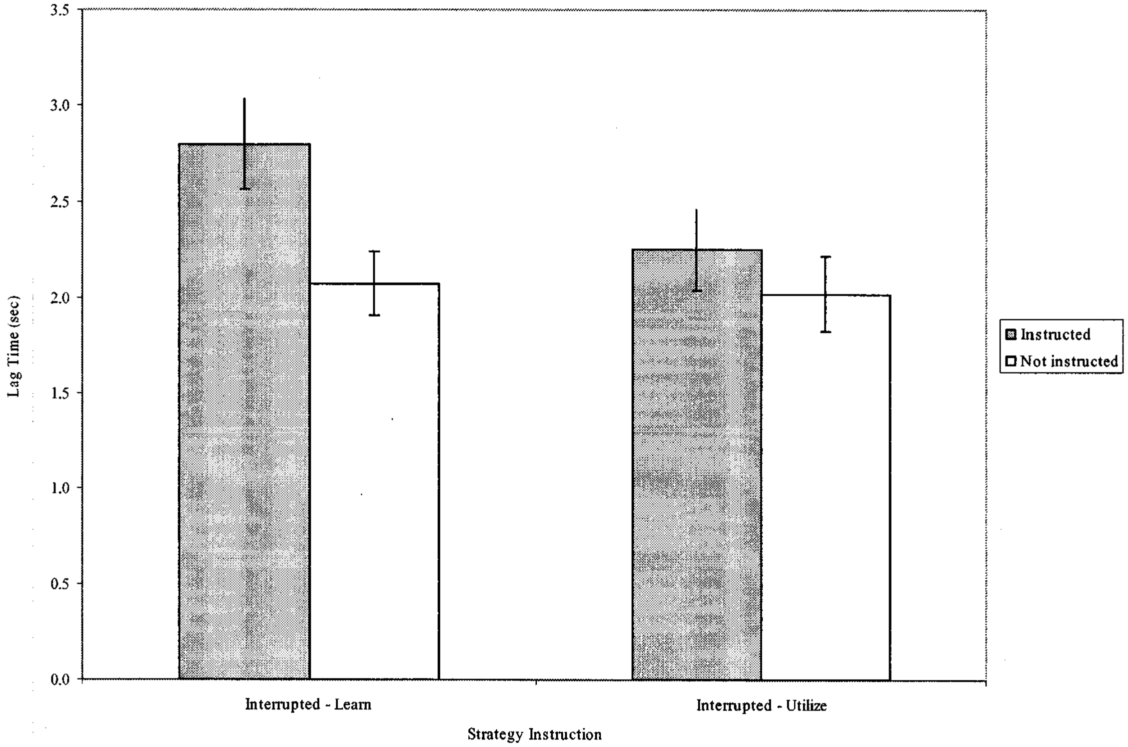


Figure 17. Average lag times were longer for instructed participants than for non-instructed participants for Interrupted-Learn aircraft only.

Were participants using the strategy as intended? In follow-up interviews, participants who received no rehearsal strategy instruction generally reported "starting over" in their decision making process following an interruption. Seventy percent or 7 of 10 uninstructed participants[1] reported that they struggled to remember the task resumption point after processing a message and thus had to start over when making a decision. One participant reported counting on fingers to remember the task resumption point. Only two participants ever mentioned using their memories to improve task resumption, reporting they "just remembered what (they were) doing."

What kinds of strategies did instructed participants use? In contrast to the uninstructed participants, only 30% (3 of 10) instructed participants reported starting over much of the time because it

---

[1] Two participants in each group were too vague in their responses to definitively determine any strategy. Thus, only 10 participants in each condition are considered here.

was difficult to remember the task resumption point. However, only 2 participants reported using the rehearsal strategy as instructed. One used a combination of the rehearsal strategy and the start over strategy. The remaining four participants reported using their fingers to remember the task resumption point.

Comparing these reported strategies indicates that most uninstructed participants started over following an interruption (70%) compared to instructed participants (30%). However, most instructed participants (at least 60%) made an active effort to remember the task resumption point, either by using physical reminders (e.g., fingers) or by rehearsing as instructed. Therefore, instruction, although modified by participants, was successful in encouraging people to use the interruption lag to remember where they left off in the interrupted task. This is borne out by the longer interruption lags of instructed participants for Interrupted-Learn aircraft compared to the interruption lags of uninstructed participants.

However, this finding does not explain why the instructed participants' decisions were more disrupted by interruptions. In order to understand this effect, participant self reports were examined in more detail. All participants, regardless of condition, were asked why they "started over" following an interruption. Why did they not try to remember the task resumption point? Participant responses focused on the notion that remembering the task resumption point "didn't matter." They were aware that after reading a message, all the interrupted task information would be available on-screen. However, the message information would be lost if it was not remembered. Thus, forgetting the task resumption point was less consequential than forgetting the message information. Therefore, it appears that participants made a conscious trade-off between the effort associated with remembering the message and the effort associated with remembering the task resumption point.

This emphasis on remembering the message may explain why rehearsal strategy instruction was disruptive to interrupted decisions. Participants may have focused on encoding the message to the extent that it interfered with memory for task resumption. Specifically, memory for message content, which was emphasized among participants, may have interfered with memory for resuming interrupted decisions making. Participants who did not make an effort to encode the task resumption point actually had an easier time resuming interrupted decisions.

In conclusion, interruption was found to be disruptive to the time it takes to make a decision. Consistent with previous research (e.g., Gillie & Broadbent, 1989), interruption adds to decision time because it requires a person to divert attention from the main task and focus on the interruption. This results in a resumption lag which is part of this overall decision time. Interestingly, however this finding did not hold for interruptions that were immediately relevant. People were fairly adept at integrating interruptions with task resumption if the interrupting message was immediately relevant. This is an unexpected finding and a unique contribution to the research literature. It has important implications for designing how and when people should be interrupted. Interruptions that people can immediately integrate into their on-going performance take no longer than non-interrupted decision and possible boost decision accuracy.

Experiment 1 was also unique because in addition to interruption effects on the interrupted task, the lingering effects of interruptions were examined. Would interruptions affect not only the interrupted decision but subsequent decision making? Interruptions did not result in a deterioration of subsequent decision-making. Neither Memory Load aircraft nor Recall Message aircraft resulted in worse performance compared to baseline decisions. Interestingly, Recall Message decisions were actually faster than baseline decisions. As described above, interruptions were shown to negatively affect the interrupted decision. However, people compensated for this by completing a part of the later decision process during the interruption. They converted the raw message data to a simpler rating value during the interruption rather than later when the relevant aircraft was on the screen. This saved them time and resulted in faster decision-making for Recall Message aircraft.

In addition to examining the on-going effects of interruption across entire task performance, Experiment 1 was designed to see if people can successfully use their memories to alleviate the disruption caused to interrupted tasks. Instruction in the rehearsal strategy was expected to be a modest way to alleviate the disruption evidenced in Interrupted-Learn aircraft. Indeed, participants in the instructed condition were found to use both memory and physical strategies to try to remember the task resumption point. However, the effort was not successful. Instructed participants actually had longer decision times than non-instructed participants. Instructing participants to use the rehearsal strategy was a conservative effort to improve performance and speed decision-making. However, using the interruption lag to remember the task resumption point did not improve performance. The strategy had a strong disruptive effect itself and compounded the effect of interruption.

Previous research overwhelmingly indicated that resuming the main task is difficult following an interruption resulting in an overall increase in time-on-task. In the experiment reported here, this effect is present as indicated by an increase in decision time for Interrupted-Learn aircraft compared to Complete aircraft. The experiment reported here sought only to try to minimize this effect by training participants in a memory strategy. However, the finings were very surprising and offered evidence that people were more influenced by the memory requirements of the interrupting task (the message) than expected. The evidence for this is the greater disruption to interrupted tasks by participants trained in the rehearsal strategy. Table 5 summarizes this finding.

Table 5. Summary of Experiment 1 findings.

		Participant Trade-offs	
	Experiment Goals	Task Resumption Point	Interrupting Task
1	The interruption lag rehearsal strategy will reduce the disruption to decision times caused by interruptions.	Training in the rehearsal staragey worsened decision time. Participant reports indicated that there was interference when trying to remember both the task resumption point and the interrupting task.	

It seems that participants made a conscious trade-off between the effort associated with remembering the message and the effort associated with remembering the task resumption point. Remember that message information was transient, but interrupted task information was permanently retained on the screen. Although instructed participants seemed to implement strategies for remembering the task resumption point, they also recognized that the cost of forgetting the message content was higher than the cost of forgetting the resumption point. They may have focused on encoding the message and this interfered with memory for the task resumption point. Non-instructed participants, who generally did not make an effort to encode the task resumption point, actually resumed the interrupted task faster. In an effort to evaluate this explanation for the unexpected findings, a second experiment was designed. It tested the trade-offs that participants make between the cost of forgetting the message and the cost of forgetting the task resumption point. It is expected that there will be evidence of both trying to remember the task resumption point and trying to remember the interrupting task message.

4. EXPERIMENT 2

The second experiment used a simplified paradigm to examine the effect of memory requirements associated with interruptions to identify evidence of both trying to remember the task resumption point and trying to remember the interrupting task message. To do this, Experiment 2 focused on the aircraft that had been most disruptive to decision time in Experiment 1, the Interrupted-Learn aircraft. Specifically, the types of aircraft examined in the second experiment are listed in Table 6.

Table 6. Types of aircraft in Experiment 2.

	Experimental Aircraft		**Control Aircraft**
Information	**Interruption**	**No Interruption**	**No Interruption**
Complete	Interrupted (Read message for future use).	Memory Load (Remember message for future use).	Complete (Baseline best decision).
Missing		Recall Message (Use message read earlier).	

The cost of forgetting the task resumption point was manipulated by adjusting interruption timing. Half the interruptions came early in the decision process (3 seconds after an aircraft appeared), arriving when only a small amount of decision work had been completed. Thus, the cost of forgetting the task resumption point was very small, i.e., starting decision-making over is not very effortful or time consuming. In contrast, half of the interruptions came late in the decision process (7 seconds after an aircraft appeared), arriving when a larger amount of decision work had been completed (i.e., a decision was near completion), and the cost of forgetting the task resumption point is very high.

The cost of forgetting the message content was manipulated by adjusting the importance of the message data. For this experiment, all interruptions were relevant to a future aircraft (three aircraft ahead). One scenario was designed so that messages contained data that were unnecessary for a decision maker to remember. Specifically, the message data were redundant. In such a scenario, aircraft never had missing information. Such messages served to interrupt, but were purely distracting, because participants did not need to remember the message content. Thus, forgetting these messages had a low cost. In contrast, another scenario was designed so that messages contained data that was necessary for a decision maker to remember, as in the first experiment. In this scenario, aircraft do occasionally have missing information requiring a participant to remember message content. Forgetting the content of these messages had a high cost.

Table 7. Interruptions varied in the cost of forgetting associated with message type (unnecessary, necessary) and message timing (early, late).

			Message	
			Low Cost	**High Cost**
			Message unnecessary	Message necessary
Task Resumption Point	**Low Cost**	Decision interrupted early (3 seconds)	*Remember neither*	*Remember message*
	High Cost	Decision interrupted late (7 seconds)	*Remember resumption point*	*Remember both*

Table 7 depicts the characteristics of interruptions in the second experiment. Each of the four resulting cells reflects the different ways in which people might process an interruption. When the message is necessary and arrives early, people are expected to focus on remembering the message. This should be exhibited by longer message times. In contrast, when the message is unnecessary and arrives late, people are expected to focus on the remembering the task resumption point. This should be exhibited by longer interruption lag times. When both the message and task resumption point must be remembered, decisions are expected to be most disrupted as people try to balance the requirement to remember both the message and the task resumption point. Disruption is expected to be least when the cost of forgetting the message and the task resumption point are both low and neither needs to be remembered.

METHOD

Participants

Ten undergraduates volunteered to participate for two-hours in exchange for course credit. Seven men and three women participated, and their median age was 22.00 years (SD=2.36). All participants received messages containing altitude information, although the necessity of this information was differed between the two scenarios presented (as described above). Each participant was assigned to one counterbalanced condition to determine which type of scenario would be viewed first, either necessary messages or unnecessary messages.

Experiment Procedure

Participants' instruction was identical to the procedures in Experiment 1. However, all of the participants were instructed in the rehearsal strategy in an effort to create a situation similar to Experiment 1 in which there significant competition is hypothesized between remembering the task resumption point and the interrupting message. When each participant reached criterion (7/10 correct) on decision-making training, he or she was allowed to move on to the message-specific training. Participants were assigned to receive first either a necessary or unnecessary message scenario. Following initial training in decision-making, each participant received training for the first scenario. Participants were given no instructions other than that they would be reading messages in addition to making decisions. When each participant reached criterion (7/10 correct), he or she was told that the experimental scenario were be the same as the practice trial. Then the participant was given a break and allowed to move on to the appropriate experimental scenario. Following the first experimental scenario, this procedure was repeated for the second scenario.

In each of the two scenarios, regardless of whether messages were necessary or unnecessary, half of the messages arrived early (3 seconds after the aircraft appeared) and half of the messages arrived late (7 seconds after the aircraft appeared). Differences in message timing were randomly presented.

After completing the second experimental scenario, each participant was asked a series of questions about the task. Then, the participant was debriefed.

RESULTS AND DISCUSSION

Figure 18 provides an overview of differences in decision-making processes for interrupted and non-interrupted aircraft in Experiment 2.

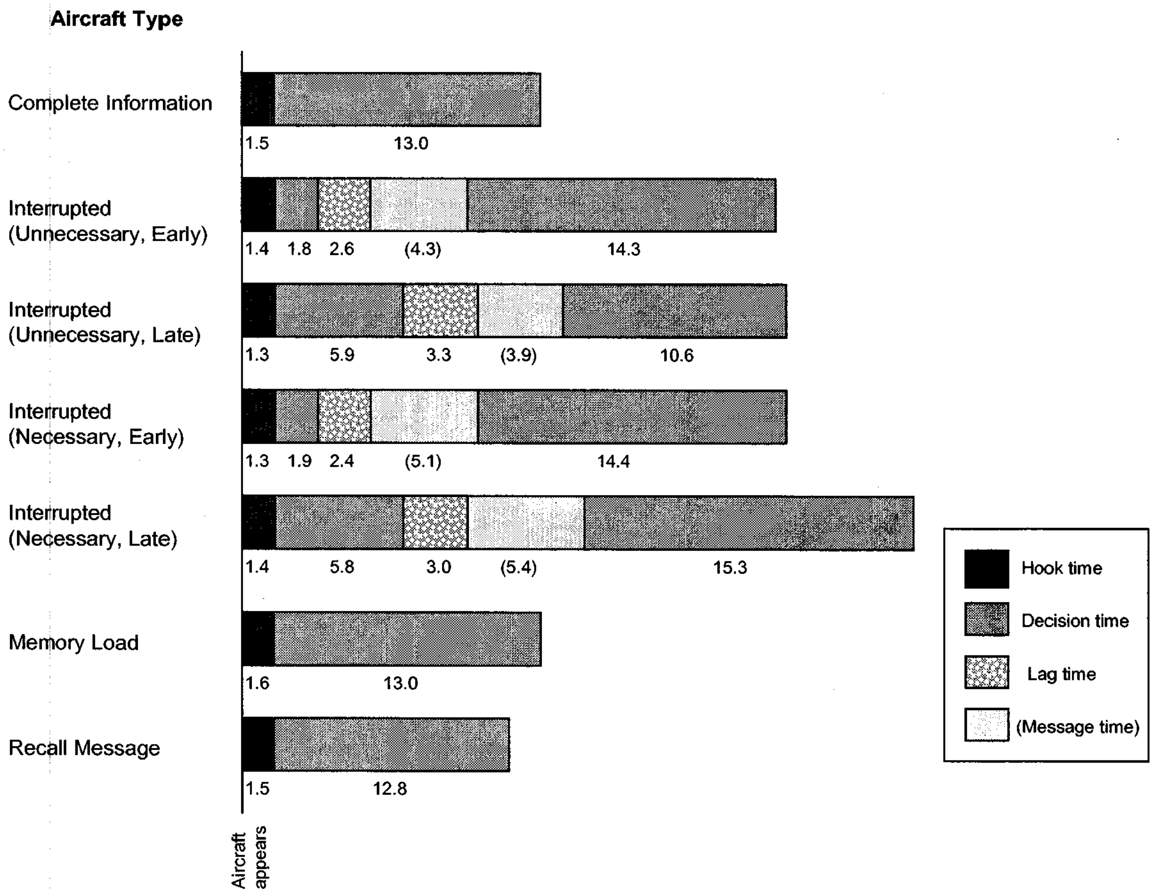


Figure 18. Time-on-task for different aircraft types in Experiment 2.

Experiment 1 had suggested that people were making unexpected trade-offs between trying to remember the task resumption point and trying to remember the interrupting task message. Recall that the goal of Experiment 2 was to distinctly provide evidence of each of these by examining a) the conditions under which decision time would be most degraded by interruptions and b) the relative interruption lag time and message time for these conditions.

How disruptive to decision time were the four types of interruption shown in Table 7? There was a significant main effect of timing, F(1,8)=8.336, p=0.020 and a marginal effect of message type, F(1, 8)=3.757, p=0.089. There was also a marginal interaction between message type and timing, F(1,8)=4.886, p=0.058, shown in Figure 19. Messages that were necessary and arrived late (7 seconds) had marginally longer decision times (M=21.09, SD=6.71) compared to all other scenario aircraft (M=16.27, SD=1.26), HSD=4.09 These results indicate that there was a significant amount of disruption when the costs of forgetting were highest (messages were necessary but arrived late after much decision processing had been completed). This finding is consistent with the expectation that remembering both the message and the task resumption point would be most disruptive to decision making.

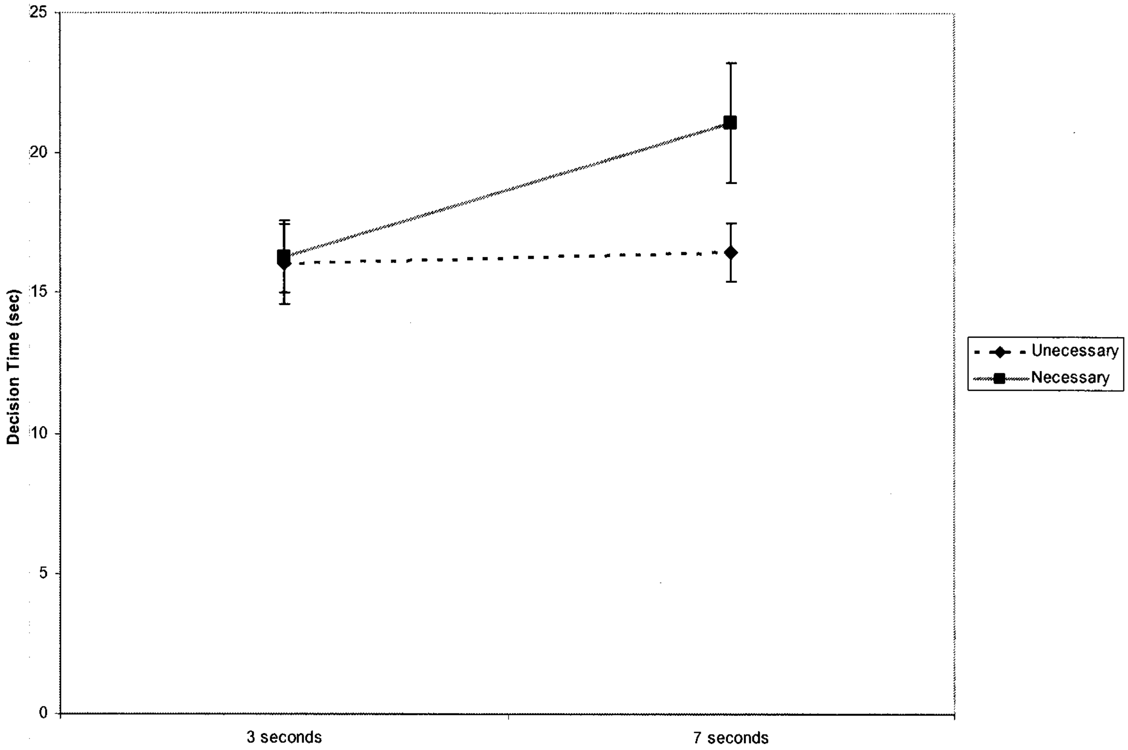


Figure 19. Decision time was longest when the costs of forgetting were highest (messages were necessary and arrived late).

Did participants have longer interruption lags or message times based on the characteristics of the interruption? These findings would offer evidence of the trade-offs suggested in Experiment 1. If people valued future decision accuracy, remembering the message would be emphasized and message times would be long. If people valued minimizing task disruption, remembering the task resumption point would be valued and interruption lags would be long.

Examining message time, there was no main effect of interruption timing, F(1,9)=0.242, p=0.635, nor an interaction of type and timing, F(1,9)=0.643, p=0.443. However, there was an effect of message type that approached significance, F(1,9)=3.982, p=0.077. When messages were necessary, participants spent one second extra reading the message compared to when messages were unnecessary. Average message time for necessary messages was 5.2 seconds (SD=0.75), and average message time for unnecessary messages was 4.2 seconds (SD=0.70). These times are consistent with the expectation that people place value on remembering messages that are necessary.

It is surprising that participants spent any time at all reading messages that were unnecessary. Examination of participants' self-reports indicated that participants might have doubted the stability of information available from messages. Three of the ten participants reported that they simply dismissed the unnecessary messages and did not try to remember them. However, three participants reported reading the unnecessary messages "just in case," "to make sure it was correct," and "to make sure it didn't change." The remaining participants did not clearly report their approach to unnecessary messages. The analysis seems to show that participants were sensitive to the fact that unnecessary messages were in fact unnecessary as indicated by the shorter message times. However, participants did not entirely trust that the experimenter wasn't going to manipulate the message content. At least some of the participants paid attention to unnecessary messages, suggesting that training should have more strongly emphasized message information would not change.

Prior to attending to these messages, how much time did participants spend in the interruption lag? There was no main effect of type, F(1,9)=0.381, p=0.552, nor an interaction of type and timing, F(1,9)=0.019, p=0.894. However, the effect of message timing approached significance, F(1,9)=4.399, p=0.065. When messages arrived late in the decision-making process, participants appeared to spend extra time during the interruption lag compared to when messages arrived early. Average interruption lag time for late-arriving messages was 3.12 seconds (SD=1.09); for early-arriving messages it was 2.50 seconds (SD=0.47).

The results thus far show that people adjust their behavior based on message characteristics. Participants spend more time reading necessary messages than unnecessary messages, and they spend more time in the interruption lag when messages arrive late rather than early. These behaviors reflect trade-offs that participants made based on the cost of forgetting the message or the task resumption point, respectively. When participants needed to remember both the message and the task resumption point, disruption to the interrupted task was greatest.

The first experiment revealed the unique ways that interruptions can disrupt performance in a realistic task, one in which interruptions are frequent and necessary. It also suggested that trying to process these interruptions is a tricky balancing act because people must remember both the interrupting message and the task resumption point in order to succeed. The second experiment tested the hypothesis that people make processing trade-offs based on the memory requirements of the task. The results indicate that participants made strategic trade-offs in the expected way. Participants spent more time reading message interruptions that were necessary than those that were unnecessary, and they had longer interruption lags when messages arrived late in the decision process than when they arrived early. This indicates that participants were sensitive to the relative costs of forgetting the message and the task resumption point. Table 8 highlights these findings so far.

Table 8. Evidence from Experiments 1 and 2 of the trade-offs made by participants in balancing the requirements to remember the task resumption point and the interruption itself.

	Experiment Goals	Participant Trade-offs	
		Task Resumption Point	Interrupting Task
1	The interruption lag rehearsal strategy will reduce the disruption to decision times caused by interruptions.	Training in the rehearsal staragey worsened decision time. Participant reports indicated that there was interference when trying to remember both the task resumption point and the interrupting task.	
2	Characteristics of the interrupted task (interrupted early or late) and the interruption itself (necessary or unecessary) will determine a) how disruptive to decision time the interruption is and b) how people process the interruption.	Decision time was longest when messages were necessary and arrived late.	
		Interruption lag time was longest when messages arrived late.	Message time was longest when messages were necessary.

These findings are evidence that people are capable of juggling different types of interruptions by strategically adjusting the time they spend attending to the task resumption point and to the interrupting message. Under what conditions might these findings change? A third experiment was designed so that the consequences of forgetting the task resumption point might be more calamitous. Experiment 3 replicated the design of the second experiment, but the complexity of the decision task was increased. It was expected that forgetting the task resumption point would be more consequential because the process of making a decision was more laborious. Thus, people were expected to emphasize remembering the task resumption point over remembering the message.

5. EXPERIMENT 3

The purpose of the third experiment was to evaluate the influence of task complexity on interruption processing trade-offs. Task complexity was increased by modifying the complexity of the decision rule participants used to make decisions. As in the second experiment, it was expected that people would behave differently depending on the costs of forgetting the message or the task resumption point. However, the more laborious decision task is expected to make timely task resumption more valuable than in the last experiment. Specifically, when interruptions arrive late in the decision process, it is expected that decision time will be greatest because resuming the main task will be most difficult.

METHOD

Participants

Ten undergraduates volunteered to participate for two-hours in exchange for course credit. Five men and 5 women participated, and their median age was 20.00 years (SD=3.09).

Experiment Procedure

The procedures for this experiment were identical to that of the second experiment. However, in order to increase task complexity, the decision rule in which participants were instructed differed from the simple additive rule that was used in the previous two experiments. Appendix E provides this rule which included multiplicative terms for each cue in the equation. This increased the complexity of the decision task by requiring participants to perform more calculations and hold more interim terms in memory.

RESULTS AND DISCUSSION

Figure 20 provides an overview of differences in decision-making processes for interrupted and non-interrupted aircraft in Experiment 3.

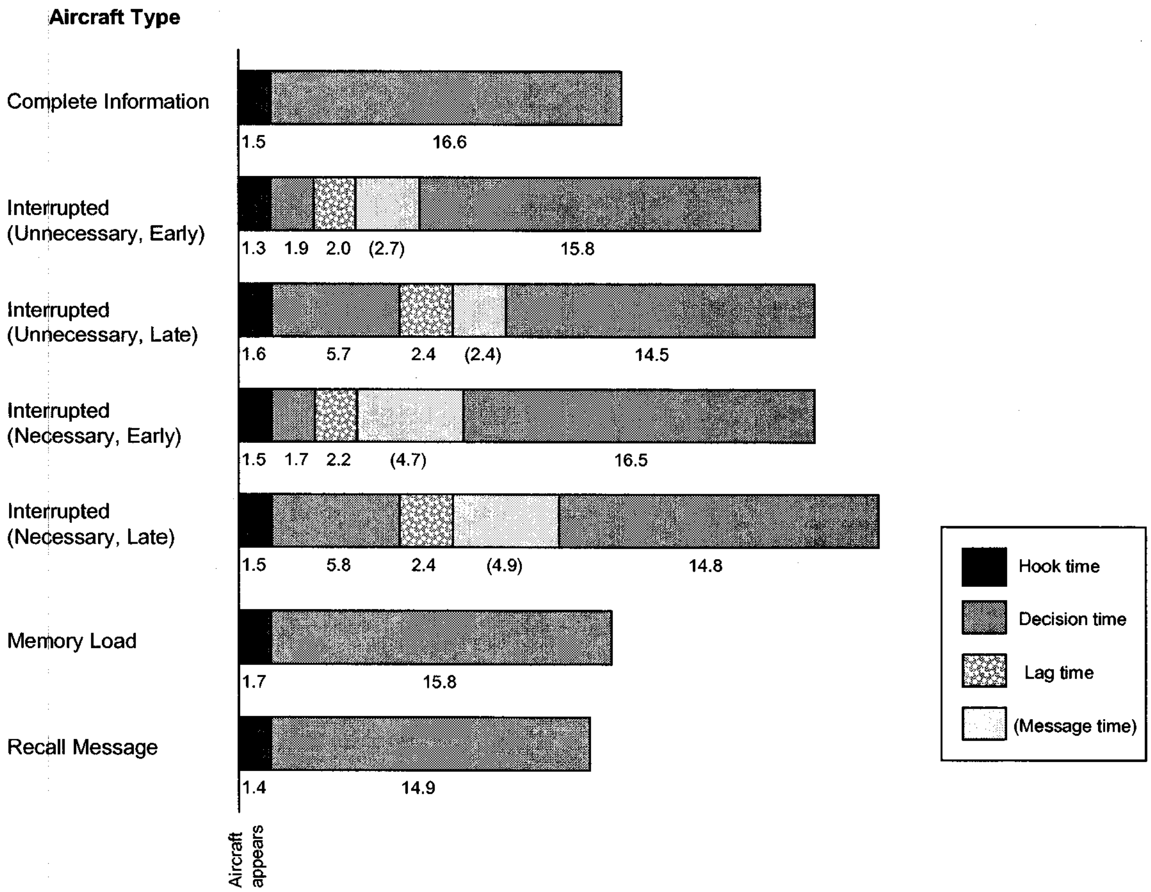


Figure 20. Time-on-task for different aircraft types in Experiment 3.

Overall, decision times were longer in the third experiment (e.g., Interrupted 19.09, Complete 16.60 seconds) than in the second experiment (e.g., Interrupted 17.48, Complete 12.98 seconds). Here, decision time for interrupted aircraft varied in marginal ways as a result of message timing, F(1,9)=4.22, p=0.070. There was no effect of message type, F(1,9)=0.064. p=0.806, nor was there an interaction of F(1,9)=0.004, p=0.948. These results are partially consistent with those in the second experiment. There, interruption with necessary messages was more disruptive when messages arrived late than when messages arrived early. However, in the current experiment, this effect is present regardless of whether interruptions were necessary or unnecessary. This finding is likely the result of the increased decision complexity.

Here, even the minimal distraction required to dismiss an unnecessary message interferes with task resumption.

Unnecessary messages seemed as disruptive as necessary messages. However, participants spent less time reading unnecessary messages. In the second experiment, when messages were necessary, participants spent one second extra reading the message compared to when messages were unnecessary. Here again, when messages were unnecessary, there was a significant main effect of message type, F(1,9)=6.781, p=0.029. In this experiment, there was no effect of message timing, F(1,9)=0.027, p=0.873, nor was there an interaction of timing with message type, F(1,9)=0.769, p=0.403. Participants spent an average of 2.6 seconds (SD=1.76) reading unnecessary messages compared to 4.8 seconds (SD=3.70) when messages were necessary. This is a difference of 2.2 seconds longer for necessary interruptions. This effect is in the same direction as in the second experiment. However, message times were longer in the last experiment, and the difference between message types was smaller. In the second experiment, participants spent on average 4.2 seconds reading unnecessary messages. In the current experiment, this time was cut by half. It is possible that the increased decision-making complexity of the third experiment led participants to focus more on the main interrupted task and thus to devote less time to reading messages.

If people tended to focus more on the interrupted task (and remembering the task resumption point), it is expected that they would adjust the time they spend in the interruption lag. In the second experiment, participants spent more time in the interruption lag when messages arrived late compared to when messages arrived early. In this experiment, there was a marginally significant main effect of message timing, F(1,9)=4.076, p=0.074. When messages arrived early, participants spent 2.1 seconds (SD=0.44) in the interruption lag compared to 2.4 seconds (SD=0.75) when messages arrived late. These times support the trend that people are using the interruption lag to prepare to process the interruption and resume the main task. However, the small size of this effect was surprising since resumption of a complex task was expected to be very challenging. A possible explanation is that this effect would have been larger if late interruptions had occurred later than 7 seconds into the decision task. Because decision-making takes longer in the complex task of Experiment 3, an interruption at 7 seconds may be less costly than in the simpler task of Experiment 2.

The results of this third experiment support the idea that people strategically balance the memory requirements associated with interruptions. Consistent with Experiment 2, participants spent more time reading necessary messages than unnecessary messages. They also spent marginally less time in the interruption lag when messages arrived early compared to when messages arrived late. However, did people emphasize task resumption more than in the previous study? This was expected due to the increased complexity of the decision rule and the increased costs associated with forgetting the task resumption point. The results show that later messages were more disruptive than early messages regardless of whether they were necessary or not. The more complex nature of the decision task created a situation where any message disrupted performance, interfering with task resumption. Table 9 summarizes this finding in the context of Experiments 1 and 2.

Table 9. Evidence from Experiments 3 indicates that participants made slightly different trade-offs when task complexity was greater than in previous experiments.

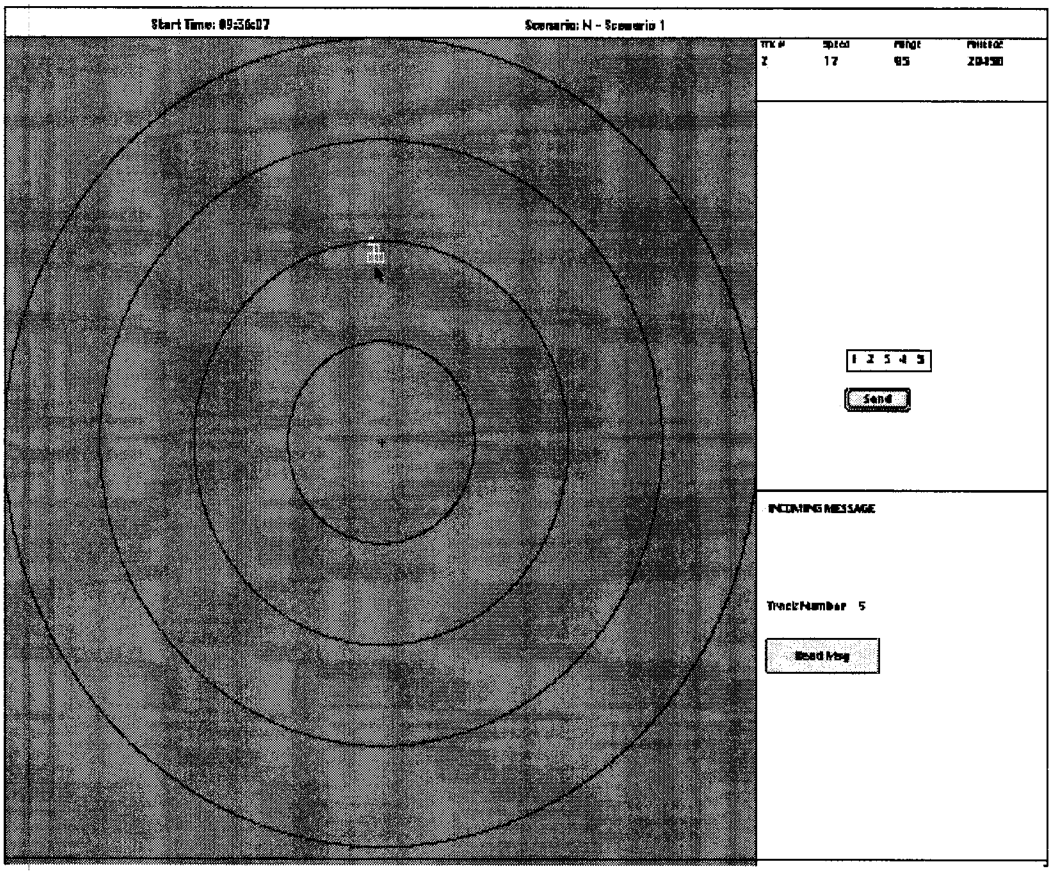
		Participant Trade-offs	
	Experiment Goals	Task Resumption Point	Interrupting Task
1	The interruption lag rehearsal strategy will reduce the disruption to decision times caused by interruptions.	Training in the rehearsal staragey worsened decision time. Participant reports indicated that there was interference when trying to remember both the task resumption point and the interrupting task.	
2	Characteristics of the interrupted task (interrupted early or late) and the interruption itself (necessary or unecessary) will determine a) how disruptive to decision time the interruption is and b) how people process the interruption.	Decision time was longest when messages were necessary and arrived late.	
		Interruption lag time was longest when messages arrived late.	Message time was longest when messages were necessary.
3	When the interrupted task is sufficiently complex, task resumption will be more difficult as indicated when interruptions arrive late.	Decision time was longest when messages arrived late.	
		Interruption lag time was longest when messages arrived late.	Message time was longest when messages were necessary.

The fact that participants continue to read unnecessary messages was troubling. However, this may have been an artifact of the experimental system being used. Participants needed to, at a minimum, click a series of buttons to dismiss a message even if they wanted to ignore the message. Thus, interruption was inherent in the task, even when the interruption content, the message, was unnecessary. However, in some real world tasks, the interface often plays a smaller role in the processing of interruptions. A fourth experiment was designed to further explore this effect. In the fourth study, interrupting messages require a motor response (clicking buttons) only if the message is to be read. If a participant chooses to ignore a message, he or she needs only to continue the main task. In addition to being a realistic alternative in real word decision-making, this modification allows an examination of true ignoring.

6. EXPERIMENT 4

The first three experiments were used to examine the way that people process interruptions, particularly with regard to balancing the need to remember the interrupting message and the need to remember the task resumption point. So far, in each of these studies, participants have been required to physically react to a message alert and the message itself. Specifically, each participant had to click the "Read" and "Done" buttons before switching attention back to the interrupted task. Experiment 4 was designed to investigate the situation where the message alert required no physical response. Here, a participant could click the "Read" button or simply continue working on the main task and ignore the message alert.

Procedurally, this experiment replicates Experiment 2 using a simple decision-making rule, but it uses a modified interface in which participants could ignore or "not read" an interrupting message. Figure 21 shows the I-Argus interface displaying a message alert in the lower right-hand side of the screen. In previous experiments, this alert was displayed in a window that covered the radar scope, preventing continued work on the main task. In this fourth experiment, participants could press the "Read" button or ignore the alert and continue to attend to the main task.



Start Time: 09:36:07

Scenario: N - Scenario 1

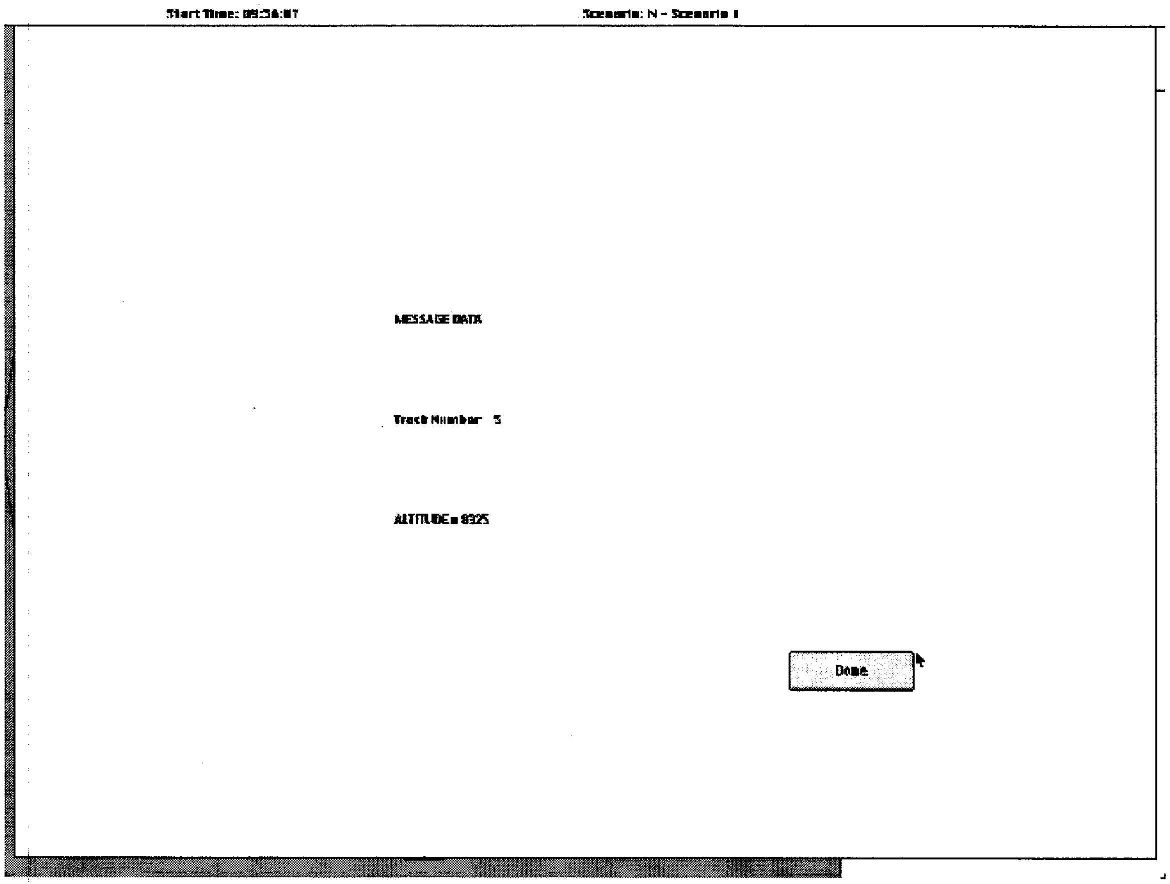
TRK #	Speed	Range	Altitude
2	17	95	20450

INCOMING MESSAGE

Track Number 5

Figure 21. The I-Argus interface used in Experiment 4.

Upon pressing the "Read" button, the message appears on the screen, as shown in Figure 22. If the "Read" button is not pressed, than the alert will disappear from the screen after 5 seconds.



Start Time: 09:56:07

Scenario: N - Scenario 1

MESSAGE DATA

Track Number 5

ALTITUDE = 8325

Figure 22. The message appears on the screen after pressing the read button in all four experiments.

METHOD

Participants

Twelve undergraduates volunteered to participate individually in three hour sessions. Six of these were recruited during the summer semester and were paid an honorarium of $20.00. Six were recruited during the fall semester and participated in exchange for course credit. A total of 8 men and 4 women participated, and their median age was 20.50 years (SD=1.90).

Experimental Procedure

The procedures for this experiment were identical to that of the second experiment in which the simple additive rule was used. However, there were two procedural modifications. First, each participant was trained to use the modified interface, rather than the original interface used in the first three experiments. Second, none of the participants was instructed in the rehearsal strategy.

When each participant reached criterion (7/10 correct) on decision-making training, he or she was allowed to move on to the message-specific training. Participants were assigned to receive first either a necessary or unnecessary message scenario. Following initial training in decision-making, each participant received training for the first scenario. Participants were given no instructions other than that they would now be reading messages in addition to making decisions. They were assured that all data was accurate; they would never be given any false or misleading information. When each participant reached criterion (7/10 correct), he or she was queried as to their understanding of the scenario. Participants were expected to understand that messages were either necessary or unnecessary (as determined by the scenario presented). When the experimenter felt assured each participant understood this, the participant was given a break and allowed to move on to the appropriate experimental scenario. Following the first experimental scenario, this procedure was repeated for the second scenario.

RESULTS AND DISCUSSION

Figure 23 provides an overview of differences in decision-making processes for interrupted and non-interrupted aircraft in Experiment 4. This figure shows interruption lag times and message times for the aircraft messages that were read. However, a unique aspect of the fourth experiment is that participants had several alternatives in how they reacted to a message. As in the previous experiments, they could decide to read the message when the alert appeared. However, they also could finish making a decision before reading the message. Finally, participants could also choose to not read the message at all, essentially ignoring the interruption. This analysis will include a comparison of the proportion of times each of these options is exercised by a participant.

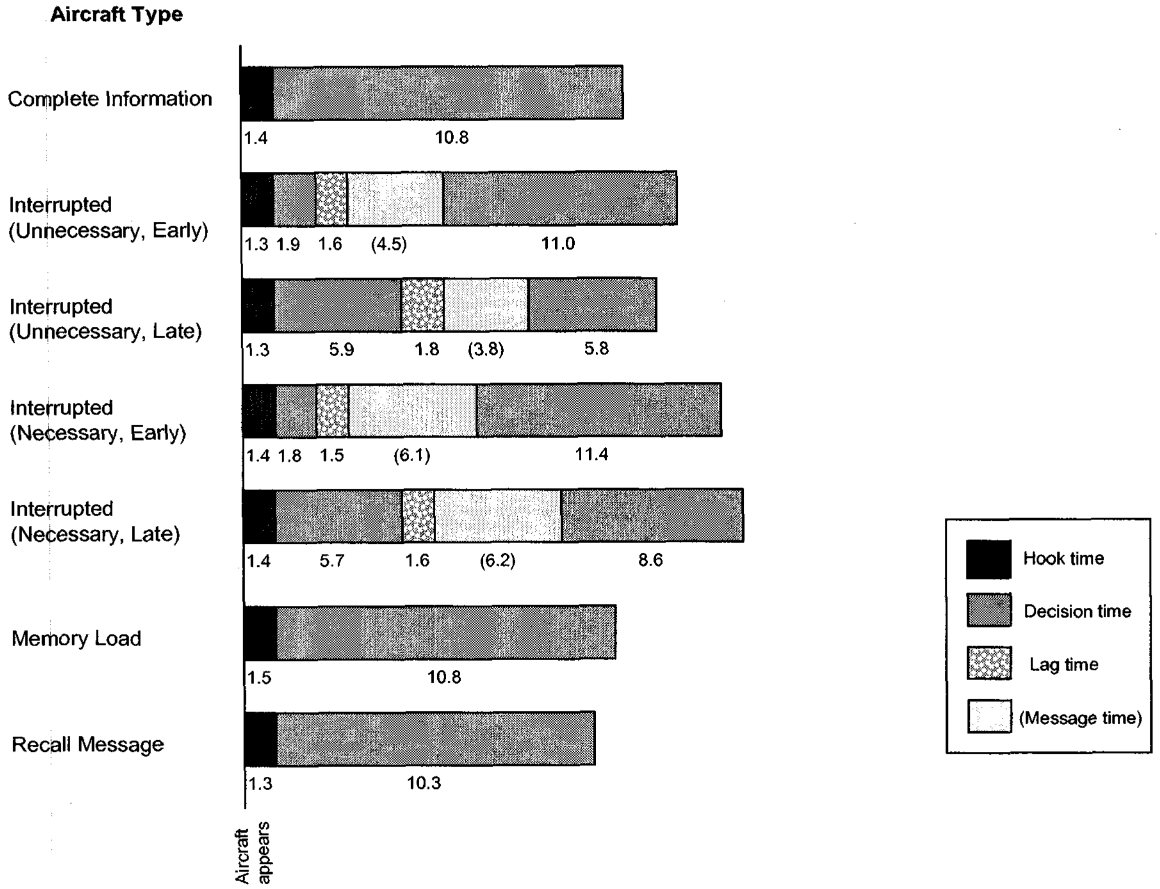


Figure 23. Time-on-task for different aircraft types in Experiment 4.

Decision times did not vary as a result of differences in message timing, F(1,11)=0.033, p=0.858 or message type, F(1, 11)=1.529, p=0.242. However, there was a marginal interaction of message type and timing, F(1,11)=3.862, p=0.075, indicated in Figure 24. Decision times for messages that were necessary and arrived late had marginally longer decision times (M=14.36, SD=5.12) than messages that were unnecessary and arrived late (M=11.63, SD=4.28), HSD=2.74. This is very similar to the finding of Experiment 2, where necessary late messages resulted in the longest decision times. However, this result reflects only those cases for which messages were actually read.

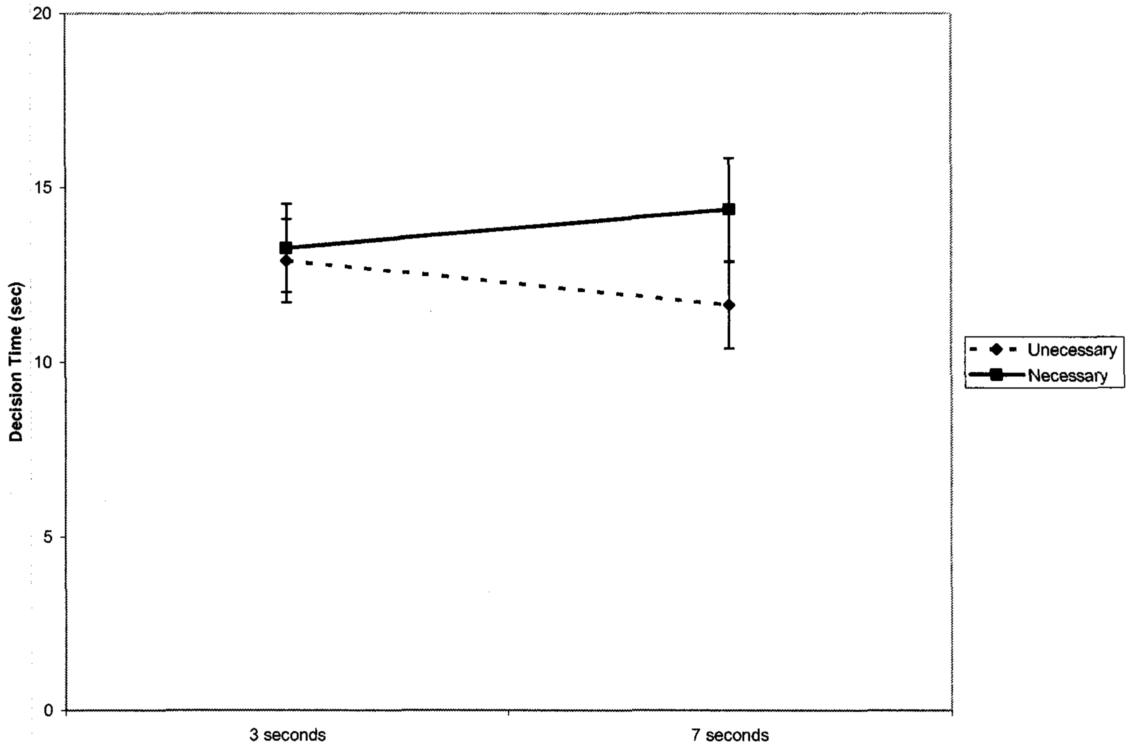


Figure 24. When messages were late, those that were necessary resulted in longer decision times than those that were unnecessary.

Did interruption type (necessary or unnecessary) or timing (early or late) affect if and when a message was read? Figure 25 shows the participants' message reading behavior across necessary and unnecessary message scenarios. There was a significant effect of interruption type, F(3,8)=5.78, p=0.021 for both the number read before issuing a decision and the number of messages not read. Participants read almost all necessary messages prior to completing a decision that had been interrupted (proportion read M=0.91, SD=0.17). However, they frequently ignored unnecessary messages (M=0.55, 0.12). Occasionally, participants read messages after completing a decision (M=0.08, SD=0.16), but this behavior was not affected by whether messages were unnecessary or unnecessary.

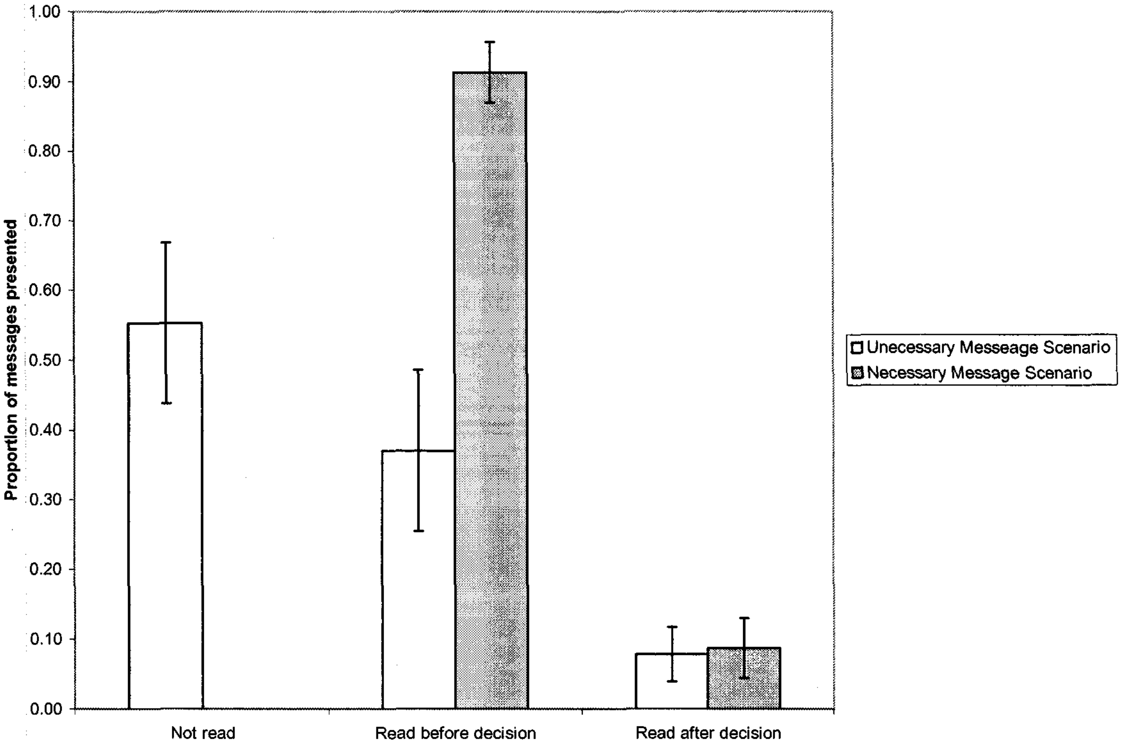


Figure 25. Interruption type affected if and when messages were read.

Although, participants showed preferential treatment of necessary messages. However, it is interesting that participants bothered to read any unnecessary messages. This behavior is similar to that observed in the previous experiments when participants spent several seconds reading unnecessary messages. Reading unnecessary messages seems to be irrational. However, this is a pervasive finding across three studies. This shows that interruptions are very enticing to decision-makers. Several explanations are possible. Participants may have become wrapped up in the rhythm of the decision task and forgotten that messages were unnecessary. Longer scenarios may have alleviated this. Alternatively, regardless of the training they received, participants may never have fully trusted the researcher and may have wanted to verify that information in messages was consistent with subsequent task information. Longer or more aggressive training may have alleviated this. Finally, this may simply be a case of participants being "unable to help themselves." The presence of unread information, even if unnecessary, may have been to tempting to ignore. Sampling from a different population (e.g., domain experts) might reduce the likelihood of this occurring.

Message time offers an indication of the amount of processing of interrupting messages that occurred. Only 6 participants were considered in this analysis, as only these six read messages before making a decision for all combinations of interruption type and timing. For messages that were read prior to a decision, there were no differences based either on interruption type, $F(1,5)=1.23$, $p=0.317$, or timing, $F(1,5)=0.145$, $p=0.719$. Although not a significant difference, when messages were unnecessary, participants spent less time reading ($M=4.18$, $SD=3.97$) than when they were necessary ($M=6.16$, $SD=2.38$). These times are very similar to those observed in the second experiment. However, many participants chose to ignore (not read) unnecessary messages, so there were fewer participants on which to evaluate the time spent reading messages. This may account for the lack of a significant finding.

Interruption lag time is how long participants delayed attending to an incoming message following the message alert. It was expected that lag time will be lower when messages arrive early compared to when messages arrive late. Again based on the 6 participants who regularly attended to a message before completing the decision, there was a marginal effect of interruption timing, $F(1,5)=4.85$, $p=0.079$. Lag time was 1.53 seconds ($SD=0.28$) for early arriving messages compared to 1.67 seconds ($SD=0.46$) for late arriving messages. Thus, as expected, participants delayed attending to late arriving messages for over one-tenth of a second. This effect is slightly smaller than in previous studies and lacks statistical significance at $p<0.05$. This is perhaps because the participants in this fourth experiment were not instructed in the rehearsal strategy. However, the trend of the result is consistent with earlier findings, indicating that people may be capable of using the interruptions lag in a meaningful way without instruction. This would be an interesting question to examine in future research.

Experiment 4 was an effort to replicate previous findings in which the interface allows the participants more options for managing interruptions. Specifically, participants in this study were able to truly ignore messages if they wanted to because the interface required no motor response to dismiss an unwanted message. The results were generally as expected in that participants frequently ignored them. However, reading some unnecessary messages continued to persist. Consistent with previous findings, interruptions continued to be disruptive to decision-making, particularly when the costs of forgetting were high (when messages were necessary and arrived late in the decision process). Table 10 summarizes the findings of Experiment 4 and the previous experiments.

Table 10. A summary of evidence from the four experiments indicating the types of trade-offs people make when processing interruptions.

		Participant Trade-offs	
	Experiment Goals	Task Resumption Point	Interrupting Task
1	The interruption lag rehearsal strategy will reduce the disruption to decision times caused by interruptions.	Training in the rehearsal staragey worsened decision time. Participant reports indicated that there was interference when trying to remember both the task resumption point and the interrupting task.	
2	Characteristics of the interrupted task (interrupted early or late) and the interruption itself (necessary or unecessary) will determine a) how disruptive to decision time the interruption is and b) how people process the interruption.	Decision time was longest when messages were necessary and arrived late.	
		Interruption lag time was longest when messages arrived late.	Message time was longest when messages were necessary.
3	When the interrupted task is sufficiently complex, task resumption will be more difficult as indicated when interruptions arrive late.	Decision time was longest when messages arrived late.	
		Interruption lag time was longest when messages arrived late.	Message time was longest when messages were necessary.
3	Participants should ignore unecessary messages when given the opportunity to do so.	Decision time was longest when necessary messages arrived late compared to early.	
		Interruption lag time was longest when messages arrived late.	Participants read almost all necessary messages, and frequently ignored unecessary messages.
			Of those that were read, message time was longest when messages were necessary.

7. CONCLUSIONS

Four experiments were conducted to investigate interruptions that are frequent and necessary. Past research focused on selected aspects of a primary or secondary task. However, no research has investigated a task were interruptions must be integrated into on-going performance. This context is important because it reflects many complicated real world situations. The primary challenge of such a context is the competing memory requirements associated with the interruption and the need to successfully resume the interrupted task. This challenge was investigated in a simulated team decision-making task.

The first experiment was used to assess the complex effects of different interruptions on initial and subsequent decisions. Interruptions were disruptive, resulting in decision times that were larger by 4 seconds when messages needed to be remembered over time. Messages that were immediately relevant to the interrupted decision task not only did not take longer but seemed to have boosted interrupted task accuracy. Even though the need to remember some messages over time causes disruption to the interrupted task, the eventual recall of message data resulted in faster decisions. Participants' reports indicate adaptation in which reading and remembering a message involved the conversion of raw data to a simpler threat value. So while interrupting messages were disruptive to decision-making, participants actually found a way to save time when recalling the earlier message.

The first experiment also examined how participants balanced remembering the message with remembering the task resumption point. In order to optimize this balancing act, half of the participants were instructed to use the interruption lag to rehearse the task resumption point. This was a conservative effort based on Altmann and Trafton's framework to minimize the disruption associated with the interruption and subsequent resumption of the task. However, the results indicate that this modest strategy actually had a negative effect on performance, increasing the time spent making decisions. Because messages were transient, they needed to be remembered or would be lost. In contrast, the task resumption point could be forgotten because there were cues on the screen to recreate the interrupted task. Participants who tried to remember both the message and the task resumption point performed worse, likely because of interference between the two items. Unexpectedly, people were attentive to unexplored task characteristics, specifically the costs associated with forgetting the message or the task resumption point. This is something that has not been explored in other research.

The next experiment tested this hypothesis by manipulating the costs of forgetting associated with the interruptions. Using a simplified paradigm, the cost of forgetting the interrupting message and the cost of forgetting the task resumption point were manipulated by varying respectively whether messages were necessary or unnecessary and whether the messages arrived early or late in the decision-making process. As predicted, participants consistently considered these characteristics in the way that they managed interruptions. Interruptions had the greatest disruptive effect on decision time when interruptions were necessary and late. Additionally, participants spent more time reading necessary messages and exhibited longer interruption lags when messages arrived late in the decision process.

The design of the third experiment increased the risks associated with interrupting the decision maker. This was accomplished by increasing the complexity of the decision task. Because decision-making was more laborious, it was hypothesized that people would adjust the way in which they balanced remembering the message and the task resumption point. The results of the third experiment differed from the second in that later messages were the most disruptive regardless of whether they were necessary or not, suggesting that the more complex nature of the decision task created a situation where any message disrupted performance, interfering with task resumption. This finding also highlighted the fact that the interface used in these experiments created unavoidable interruptions. Participants were unable to completely ignore a message.

A fourth experiment was designed so that participants could completely ignore interruptions if desired. Indeed, participants frequently ignored unnecessary messages. However, many unnecessary messages were still read, indicating the tempting nature of interruptions. Consistent with the previous experiments, interruptions continued to be disruptive to decision making, particularly when the costs of forgetting were high (when messages were necessary and arrived late in the decision process).

In summary, this series of studies revealed that interruptions have intriguing effects when they are interleaved and consequential elements of on-going performance. For one thing, interruptions that are immediately relevant don't disrupt performance. However, interruptions that have relevance for some future decision are disruptive to the interrupted task, but this disruption does not persist across future tasks (Memory Load and Recall Message aircraft). The research also established that one of the problems associated with interruptions is that people often must simultaneously remember the interruption (e.g., the message) and the task resumption point. Attempting to do so may result in interference between the two memories. Being sensitive to this, participants seem able to make trade-offs associated with the costs of forgetting either the interruption or the task resumption point. Interruptions that arrive late and are necessary are most disruptive. Participants spent longer reading necessary messages than unnecessary messages. They also spent longer in the interruption lag preparing to resume the interrupted task when messages arrived late rather than early. Finally, unnecessary messages were read for less time or ignored entirely compared to necessary messages. Reading unnecessary messages, although "irrational," was pervasive.

These findings are an interesting addition to the literature, with specific implications for Altmann and Trafton's (2003) framework. Other researchers (e.g., Gillie and Broadbent, 1989) have claimed that the opportunity to rehearse the point of interruption has not helped to minimize task disruption. The results of these experiments show consistently that people do use the interruption lag to prepare for an interruption. Further, this use is modest, lasting only milliseconds. However, people were attentive to unexplored task characteristics, specifically the costs associated with forgetting the message or the task resumption point. This is something that has not been explored in other research because other research has not occurred at this level of detail. Thus, these findings may explain why the interruption lag has received criticism as a tool for managing disruptions and may suggest ways to further study people's memory strategies as they relate to interruptions..

Although many high-level strategies have been suggested for dealing with interruptions, little research has investigated them in relation to the limitations of memory. These findings may be useful in several areas, including the design of intelligent agents, instant messaging, team communications, and vehicle telematics. System designers should strive to assess the memory requirements of not only the interrupting event but also the interrupted task and subsequent activities. For each of these, they should think about the type of cues in the world that support memory and the capabilities and limitations of a person's memory. Further research should focus in several areas. Other research using the I-Argus task could be used to manipulate the degrees of relevance and timeliness of interruptions beyond simply the necessary/unnecessary or early/late dichotomies by including information that varies in importance and the probability of being correct. Further research could also investigate whether it is possible to instruct operators in how to manage task memory requirements and process interruptions. For example, can people be trained to use the interruption lag more successfully? Also, what individual differences, untested here, contribute to interruption processing? In the first experiment, a simple working memory measure (n-back) did not predict people's use of the rehearsal strategy. Future evaluations of the trade-offs people make should consider some measure of individual memory differences, as well as decision making skill and expertise.

A complicated interruption context has not been presented in the previous literature. The findings presented here establish a basic understanding of how people process interruptions that are necessary and consequential to on-going performance particularly with respect to the memory requirements of the task. The experiments described above show people continually adapting to the memory demands of interruptions in a dynamic task. The major implication is therefore that people are active and strategic processors of interruptions.

BIBLIOGRAPHY

BIBLIOGRAPHY

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

Adams, M.J., & Pew, R.W. (1990). Situational awareness in the commercial aircraft cockpit: A cognitive perspective. In the *Proceedings of IEEE/AIAA/NASA 9th Digital Avionics Systems Conference*, 519-524.

Adelman, L., Miller, S. L., Henderson, D., & Schoelles, M. (2003). Using Brunswikian theory and a longitudinal design to study how hierarchical teams adapt to increasing levels of time pressure. *Acta Psychologica, 112*, 181-206.

Adelman, L., Miller, S. L., & Yeo, C. (2004). Testing the effectiveness of icons in supporting distributed team decision making under time pressure. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans.*

Adelman, L., Yeo, C., & Miller, S.L. (in preparation). Using Brunswikian theory and lens model equation to understand the effects of computer displays and time pressure on the performance of distributed teams. In A. Kirlick, (Ed.), *Adaptation in Human-Technology Interaction: Methods, Models, and Measures.*

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

Arroyo, E., Selker, T., & Stouffs, A. (2002). Interruptions as Multimodal Outputs: Which are the less Disruptive? In the *Proceedings of IEEE International Conference on Multimodal Interface.*

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

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

Cutrell, E. B., Czerwinski, M., & Horvitz, E. (2000). Effects of instant messaging interruptions on computing tasks. In *Extended Abstracts of CHI 2000 Conference on Human Factors in Computing Systems*.

Cutrell E., Czerwinski M., & Horvitz, E. (2001). Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In the *Proceedings of Human-Computer Interaction – INTERACT 2001*.

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

Dismukes, K., Grant, Y., & Sumwalt, R. (1998). Cockpit interruptions and distractions. *ASRS Directline Magazine*.

Dix, A., Ramduny, D., & Wilkinson, J. (1995). Interruptions, deadlines and reminders: Investigations into the flow of cooperative work. Technical Report RR9509, University of Huddersfield.

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

Field G. E. (1987). Experimentus interruptus. *ACM SIGCHI Bulletin*, *19*, 42-46.

Franke, J. L., Daniels, J. J., & McFarlane, D. C. (2002). Recovering context after interruption. In the *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*.

Fussel, S. R., Kraut, R. E., Lerch, F. J., Scherlis, W. L., McNally, M. M., & Cadiz, J. J. (1998). Coordination, overload and team performance: Effects of team coordination strategies. In the *Proceedings of CSCW 98, Conference on Computer Supported Cooperative Work.* Seattle, Washington, ACM, 275-284.

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

Hudson J. M., Christensen J., Kellogg W. A. and Erickson T. (2002). I'd be overwhelmed, but it's just one more thing to do: Availability and interruption in research management. In the *Proceedings of CHI 2002 Conference on Human Factors in Computing Systems.*

Jolly, E. J., & Reardon, R. (1985). Cognitive differentiation, automaticity, and interruptions of automatized behaviors. *Personality and Social Psychology Bulletin*, *11*, 301-314.

Kirmeyer, S. L. (1988). Coping with competing demands: Interruption and the Type A pattern. *Journal of Applied Psychology*, *73*, 621-629.

Latorella, K. A. (1996). Investigating interruptions: An example from the flightdeck. In the *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society.*

Llaneras, R. E., & Singer, J. P. (2003). In-vehicle navigation systems: Interface characteristics and industry trends. In the *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

Loukopoulos L. D., Dismukes R. K. and Barshi I. (2001). Cockpit interruptions and distractions: A line observation study. In R. Jensen, (Ed.), *Proceedings of the 11th International Symposium on Aviation Psychology.*

Mark, G., Abrams, S., & Na, N. (2003). Group-to-group distance collaboration: Examining the “space between.” In the *Proceedings of the 8th European Conference of Computer-supported Cooperative Work (ECSCW’03)*, Helsinki, Finland, September 2003.

McElree, B. (2001). Working memory and focal attention. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, *27*, 817-835.

McFarlane, D. C. (1997). *Interruptions of People in Human Computer Interaction: A General Unifying Definition of Human Computer Interruption and Taxonomy.* Washington, D.C., Naval Research Laboratory.

McKnight, J. A., & McKnight, A. S. (1993). The effects of cellular phone use upon driving attention. *Accident and Analysis Prevention*, *25*, 259-265.

Monk C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In the *Proceedings of 46th Annual Meeting of the Human Factors and Ergonomics Society*.

O'Conaill, B., & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. In the *Proceedings of CHI '95 Conference on Human Factors in Computer Systems*.

Patten, C. J. D., Kircher, A., Ostlund, J., & Nilsson, L. (in press). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis and Prevention*.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-Computer Interaction.* Reading, MA: Addison-Wesley.

Roger, D., Bull, P., & Smith, S. (1988). The development of a comprehensive system for classifying interruptions. Journal *of Language and Social Psychology*, *7*, 27-34.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.

Schoelles, M. J., & Gray, W. D. (2001). Argus: A suite of tools for research in complex cognition. *Behavior Research Methods, Instruments, & Computers*, *33*, 130-140.

Stevens, A., & Minton, R. In-vehicle distraction and fatal accidents in England and Wales. *Accident Analysis & Prevention*, *33*, 539-545.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 12,* 643-662.

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

Van Bergen, A. (1968). *Task Interruption*. Amsterdam: North-Holland.

Zeigarnik B. (1927) Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschung*, *9*, 1-85.

Zijlstra, F. R. H., Row, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, *72*, 163-185.

APPENDIX A: EXPERIMENT 1 TRAINING

Participants were instructed in the following:

- The Argus task and its similarity to real-world aircraft classification.
- The Argus team environment as a simulation of real world teams. Participants were assured that the simulation functioned like real participant teams function.
- The basic task outline. This includes how to
- Hook (select) each aircraft when it appears on the radar scope.
- Evaluate the threat associated with each piece of aircraft data (speed, range, and altitude) (Appendix A).
- Calculate how threatening (overall) the aircraft is (Appendix B).
- Determine what the proper course of action against this aircraft is, given its overall threat (Appendix C).
- Make the correct recommendation to the team leader.
- The process of receiving, reading, and dismissing messages.

This training was specific to participants' assigned strategy condition, not instructed or instructed and assigned message type, altitude or speed. All participants read:

> "It is likely that you may not always receive the information that you require in time to make a recommendation. However, please make the best recommendations you can make, EVEN WITH INCOMPLETE DATA! And remember, if your teammate sends you a message, be sure to REMEMBER THE INFORMATION that that they provide! Even if the information is not immediately relevant, it may be useful in the future!"

Half of the participants were instructed in the rehearse strategy. They read the suggestion to:

“When this screen appears, the radar scope will temporarily be blocked from view. This may initially be frustrating for you. It is difficult to be interrupted when you are making important calculations and decisions. However, real-world AWACS operators have developed a very effective strategy to manage disruptions from teammates. We suggest that you use this strategy. When you are interrupted with a message from a teammate (i.e., when the message screen appears), take a second or two to memorize the very last thing you were doing prior to the message. For example, if you had just assessed speed to be of high threat, repeat that to yourself, "speed = 3." If you had assessed the overall threat as 5, repeat that to yourself, "overall = 5." That way, when you are done processing a message, you can easily resume your prior decision task. This may seem to be a very simple strategy, but we have found it to be highly effective in making fast and accurate decisions. ”

APPENDIX B: SPEED, RANGE, AND ALTITUDE THREAT ASSESSMENT

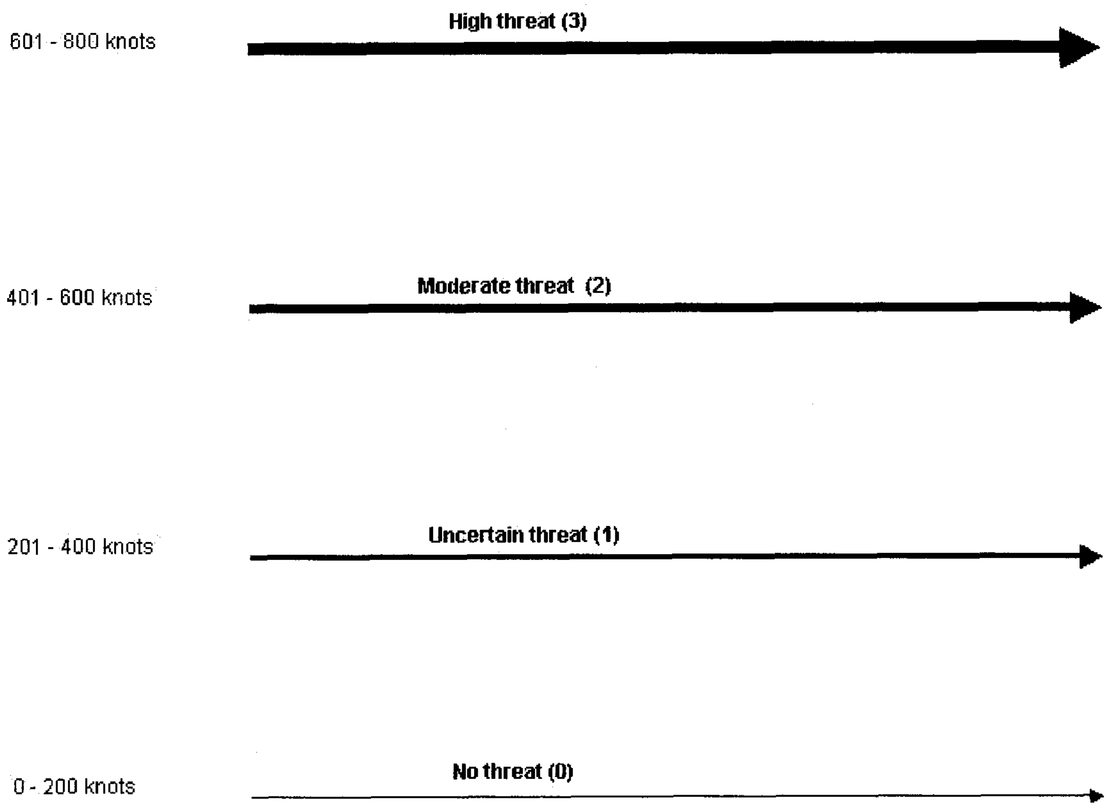
You will see raw data about each aircraft you hook, such as altitude of 1139 feet. You must then assign a threat rating from zero to three for each piece of data. These values (0, 1, 2, 3) describe the aircraft's hostility. Specifically,

0 = No threat
1 = Uncertain threat
2 = Moderate threat
3 = High threat

The following sections will describe how to transform each attribute's values into a threat rating of 0, 1, 2, or 3. You will see a chart or a picture describing each attribute. Please memorize these values and threat ratings.

Speed

This is how fast the aircraft is traveling, between 0 and 800 knots. Please memorize the following ratings:



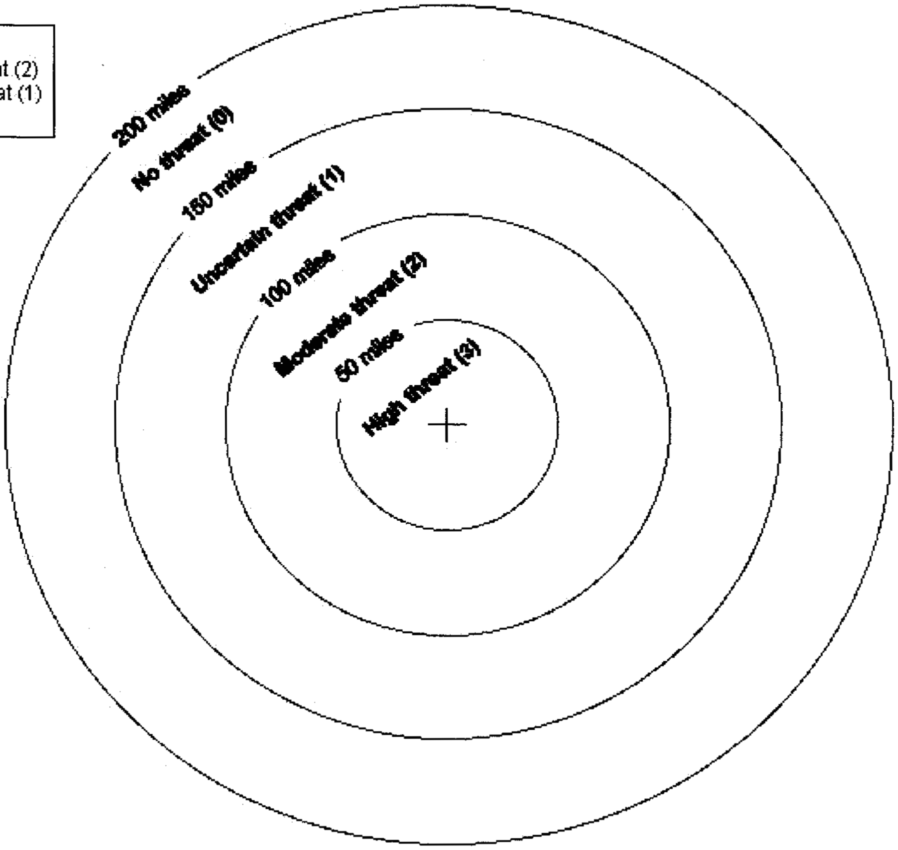
Generally, faster aircraft are more threatening than slower aircraft.

For example, a speed of 450 knots should be rated as 2 (moderate threat), but a speed of 158 knots should be rated as 0 (no threat).

Range

This is the distance between the aircraft and the Allied City, in the center of the region. Range is between 0 - 200 nautical miles. Please memorize the following ratings:

0 - 50	High threat (3)
51 - 100	Moderate threat (2)
101 - 150	Uncertain threat (1)
151 - 200	No threat (0)

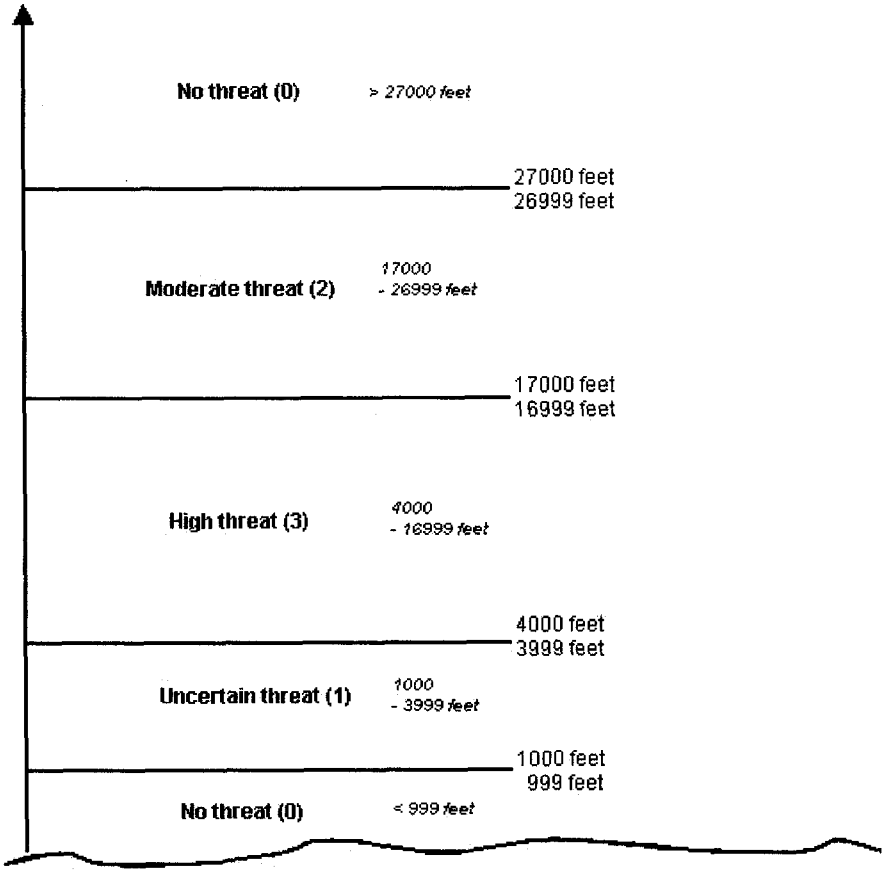


Generally, aircraft that are closer to the Allied City are more threatening than those that are farther away.

For example, a range of 129 miles should be rated as 1 (uncertain threat), but a range of 42 miles should be rated as 3 (high threat).

Altitude

This is how high the aircraft is flying from the earth. Altitude can vary between 0 - 50000 feet. Please memorize the following ratings:



Generally, aircraft that fly at intermediate altitudes are more threatening than those that fly very high or very low. For example, an altitude of 5600 feet should be rated as 3 (high threat), but an altitude of 35000 feet should be rated as 0 (no threat).

APPENDIX C: COMBINATORIAL RULE TAUGHT TO EXPERIMENT 1 PARTICIPANTS

After evaluating each piece of data for an aircraft, you need to combine the ratings to determine the aircraft's overall threat level. For example, knowing the speed of an aircraft may give an indication of how threatening it is, but by itself, that information really isn't enough to make a recommendation to shoot it down. You must use all of the attributes of an aircraft (speed, range, altitude) in order to make accurate recommendations.

There is a simple rule to determine the overall threat level of an aircraft. This rule equation provides you with a way to understand numerically how information in the world is combined to form a judgment. This rule combines all of the attribute threat ratings (0, 1, 2, 3) that are available.

$$\text{Overall Threat} = \text{Speed} + \text{Range} + \text{Altitude}$$

For example, suppose that the attribute values of an aircraft are such:

Speed is somewhat fast at 500 knots	=>	Speed ==>	2 (moderate threat)
Range is far from the city at 170 miles	=>	Range ==>	0 (no threat)
Altitude is 6000 feet	=>	Altitude ==>	3 (high threat)

Combining these terms in the rule you just studied gives us the following:

$$\text{Overall Threat} = 2 + 0 + 3 = 5$$

APPENDIX D: COURSES OF ACTION DETERMINED BY DECISION

The overall threat score allows you to choose among the courses of action available to the team. The following table lists the correct course of action based on each threat score and the definition of each course of action.

THREAT SCORE	RECOMMEND	DEFINITION
0 - 1	1 IGNORE	Recommend that the leader devote no further attention to the aircraft at the moment because it is of low threat, and instead focus on other aircraft in the area.
2 - 3	2 MONITOR	Recommend that the leader continuously track the aircraft because its status is uncertain.
4 - 5	3 WARN	Recommend that the leader send a warning to the aircraft to alter its course because it appears to be a threat.
6 - 7	4 READY	Recommend that the leader alert surface-air-missile (SAM) units toward the aircraft as a tactical position against a threat.
8 - 9	5 DEFEND	Recommend that the leader to give clearance to fire on the aircraft because attack is imminent.

In the previous example, the overall threat of an aircraft was "2 + 0 + 3 = 5." This score of 5 translates to a course of action of 3 and a recommendation to the leader to WARN the aircraft.

APPENDIX E: COMBINATORIAL RULE TAUGHT TO EXPERIMENT 3 PARTICIPANTS

After evaluating each piece of data for an aircraft, you need to combine the ratings to determine the aircraft's overall threat level. For example, knowing the speed of an aircraft may give an indication of how threatening it is, but by itself, that information really isn't enough to make a recommendation to shoot it down. You must use all of the attributes of an aircraft (speed, range, altitude) in order to make accurate recommendations.

There is a simple rule to determine the overall threat level of an aircraft. This rule equation provides you with a way to understand numerically how information in the world is combined to form a judgment. This rule combines all of the attribute threat ratings (0, 1, 2, 3) that are available.

$$\text{Overall Threat} = (2*\text{Speed}) + (1*\text{ Range}) + (5*\text{Altitude})$$

For example, suppose that the attribute values of an aircraft are such:

Speed is somewhat fast at 500 knots	=>	Speed	==>	2 (moderate threat)
Range is far from the city at 170 miles	=>	Range	==>	0 (no threat)
Altitude is 6000 feet	=>	Altitude	==>	3 (high threat)

Combining these terms in the rule you just studied gives us the following:

$$\text{Overall Threat} = (2*2) + (1*0) + (5*3)$$

$$\text{Overall Threat} = 4 + 0 + 15 = 19$$

CURRICULUM VITAE

Sheryl Miller was born on February 3, 1973, in Syosset, New York and is an American citizen. She graduated with a Bachelor of Arts in Psychology, minoring in Organizational and Occupational Behavior, from the State University of New York, College at Geneseo, New York in 1995 and a Master of Arts in Psychology, Human Factors and Applied Cognition, from George Mason University in Fairfax, Virginia in 1997.