The Pennsylvania State University
The Graduate School
College of Information Sciences and Technology

THE EFFECT OF A WORKLOAD-PREVIEW ON TASK-PRIORITIZATION AND TASK-PERFORMANCE

A Dissertation in
Information Sciences and Technology

by
Dev Minotra

© 2012 Dev Minotra

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012
The dissertation of Dev Minotra was reviewed and approved* by the following:

Michael D. McNeese  
Associate Dean for Research and Graduate Studies and Professor of Information Sciences and Technology  
Dissertation Advisor  
Chair of Committee

David L. Hall  
Dean and Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang  
Associate Professor of Information Sciences and Technology

Ling Rothrock  
Associate Professor of Industrial and Manufacturing Engineering

Madhu Reddy  
Graduate Program Director and Associate Professor of Information Sciences and Technology

*Signatures are on file in the Graduate School
ABSTRACT

With increased volume and sophistication of cyber attacks in recent years, maintaining situation awareness and effective task-prioritization strategy is critical to the task of cyber-security analysts. However, high levels of mental-workload associated with the task of cyber-security analyst’s limits their ability to prioritize tasks. Task-prioritization is especially challenging in an environment consisting of unexpected changes in relative priority between various sub-tasks and workload. Cognitive-aids that provide predictions about potential threats are designed to guide attention towards unexpected shifts in priority or workload. However, cognitive-aids may not necessarily facilitate performance under high time-pressure. They may contribute to cognitive-load resulting from display complexity and information processing demand. The literature-review explores this issue from a Human Factors perspective, particularly taking into account previous work on attention-guidance, task-management, interruptions and workload-previews. A scaled-world simulation was built to emulate cyber-security monitoring and decision-making. An experiment involving 77 participants was conducted to examine the effectiveness of a Workload-Preview under differing task-load conditions on task-performance and task-prioritization in cyber-security monitoring. Based on literature, it was hypothesized that the presence of the Workload-Preview would facilitate Task-Performance and Task-Prioritization, especially under High Task-Load. Interestingly, experimental results do not support the hypotheses. Moreover, the Workload-Preview degraded Task-Prioritization of unexpected surges under High Task-Load. This appears to have been associated to the Workload-Preview contributing to increased mental-workload. The results of the study provide implications on factors that influence the effectiveness of cognitive-aids aimed at guiding attention and improving task-performance.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. vi

LIST OF TABLES ................................................................. vii

ACKNOWLEDGEMENTS ............................................................ ix

Chapter 1 Introduction ............................................................ 1
  Motivation: Cyber-Security Event Monitoring ............................................. 4
  Research Framework ....................................................................... 7

Chapter 2 Literature-Review on Attention, Task-Management and Workload-Previews ...... 10
  Attention Theory and Workload ................................................................. 10
    The SEEV Model ............................................................................... 12
    The Multiple-Resource Theory ............................................................... 13
    The Situation Awareness Theory ............................................................. 16
    Mental-Workload ............................................................................... 19
    Workload Effects on Attentional Processes ............................................ 24
  Interruptions ......................................................................................... 26
  Task-Management, Workload and Cognitive-Tunneling ............................ 30
    Workload Effects on Task-Management ............................................... 39
  The Effect of Previews on Task-Performance ........................................ 40
    Discussion of Previous Literature on Previews ....................................... 47
    Integration and Hypotheses ................................................................ 49

Chapter 3 Materials and Methods ......................................................... 53
  The Cyber-Security Dual-Task Context ..................................................... 53
  Task-Prioritization, Event Magnitudes and Event Priority-Levels ................. 57
  Design of the Task-State Overview ......................................................... 60
  Experimental Design ........................................................................... 63
    Implementation of Independent-Variables .............................................. 64
    Implementation of Dependent-Variables ............................................... 66
  Scenario Design and Event Design .......................................................... 67
  Procedure ............................................................................................. 70
  Data Collection and Participants .............................................................. 71
  Pilot Testing ........................................................................................... 72

Chapter 4 Experimental Results ............................................................ 73
  Analysis of Main Dependent Variables ................................................... 74
    Estimation of Required Sample-Size ....................................................... 74
    Levene’s Test for Equality of Variances .................................................. 74
    Analysis of Task-Performance ............................................................... 75
    Analysis of Task-Prioritization Error ....................................................... 77
    Hypotheses Testing ............................................................................ 79
Analysis of Ancillary Experimental Variables ................................................................. 80
Analysis of TLX-Score ........................................................................................................ 81
Analysis of Surge-Prioritization Error ................................................................................ 83
Analysis of Surge Normal-Score ....................................................................................... 85
Predictors of Task-Performance ....................................................................................... 86
Summarization of Results ................................................................................................ 88

Chapter 5 Discussion ........................................................................................................ 89
Task-Prioritization and Surge-Event Effects .................................................................... 89
Effect of Task-Load ........................................................................................................... 91
Effect of Workload-Preview ............................................................................................ 92
Effects of Learning ........................................................................................................... 96
Task-Performance and Normal-Score ............................................................................... 101
Resolution on the Hypotheses ......................................................................................... 102

Chapter 6 Conclusion ....................................................................................................... 108
Implications to Human-Factors ....................................................................................... 108
Implications to Cyber-Security ......................................................................................... 112
Contributions ................................................................................................................... 113
Limitations ....................................................................................................................... 114
Future Work ....................................................................................................................... 115

References ......................................................................................................................... 116

Appendix A  Definition of Event-Categories and Priority-Levels ..................................... 129
Appendix B: Training Quiz Questions ............................................................................... 133
Appendix C: Subjective Workload Questionnaire ............................................................. 136
Appendix D: End-Task Survey ........................................................................................... 137
LIST OF FIGURES

Figure 1-1. An overview of the framework followed in this experimental study .................. 8

Figure 2-1. A Model of Human-Information Processing Stages (Wickens & Hollands, 2000) ........................................................................................................ 15

Figure 3-1. Sequence of states within an event.................................................................... 55

Figure 3-2. Screen capture of the NETS-DART interface.................................................. 56

Figure 3-3. Magnitude growth in CITIES-based simulations............................................. 59

Figure 3-4. Initial design of the Task-State Overview including the Workload-Preview .... 61

Figure 3-5. Task-State Overview component of the client interface including Workload- Preview........................................................................................................ 65

Figure 4-1. Task-Prioritization Error distribution within each performance scenario ...... 79

Figure 4-2. Frequency distribution for Surge-Prioritization Errors .................................. 84

Figure 4-3. Frequency Distributions for Surge Normal-Scores for scenarios C and D ....... 85

Figure 5-1. Preview-Availability and Task-Load interaction effect on TLX-Score .......... 93

Figure 5-2. Surge-Prioritization Error distribution for Scenario-D by Experimental condition........................................................................................................ 95

Figure 5-3. Mean Normal-Score across scenario position (between-subjects) ................. 97

Figure 5-4. Mean Task-Prioritization Errors across scenario position (between-subjects) .... 98

Figure 5-5. Mean Surge Normal-Score trend across scenario position (between-subjects) .... 98

Figure 5-6. Mean Surge-Prioritization Error trend across scenario position (between- subjects) ........................................................................................................ 99

Figure 5-7. Normal-Scores obtained on each scenario ..................................................... 101

Figure 5-8. Normal-Scores for Regular-Events and Surge-Events in Scenario D across Preview-Availability ....................................................................................... 106
LIST OF TABLES

Table 2-1. Summary of literature on task-management........................................................................39
Table 2-2. Summary of literature on workload-previews .................................................................46
Table 3-1. Mapping between event priority-levels and event magnitude-caps.................................58
Table 3-2. Translation of assigned priority-level into the number of resources applied ..........60
Table 3-3. Main Independent Variables..............................................................................................63
Table 3-4. Main Dependent Variables .................................................................................................63
Table 3-5. Additional Independent Variable.........................................................................................63
Table 3-6. Additional Dependent Variables .........................................................................................63
Table 3-7. Overview of scenario design ...............................................................................................68
Table 3-8. Minimum time required for event completion .................................................................69
Table 3-9. Examples of events in the scenario.....................................................................................70
Table 3-10. Distribution of participants by scenario ordering and experimental condition ....72
Table 4-1. Distribution of participants by academic year .................................................................73
Table 4-2. Distribution of participants by IST, SRA, CSE and other majors ...............................73
Table 4-3. Means of Normal-Scores within each performance scenario .................................75
Table 4-4. Summarization of Normal-Score differences between performance scenarios ....77
Table 4-5. Means of Task-Prioritization Error within each scenario ..............................................78
Table 4-6. Means of TLX-Score within each performance scenario ..............................................81
Table 4-7. A summarization of TLX-Score differences between scenarios .................................82
Table 4-8. Mean Surge-Prioritization Score within each condition ..............................................83
Table 4-9. Linear Regression between Normal-Score (DV) and mental-effort (predictor) ....86
Table 4-10. Linear Regression between Surge Normal-Score (DV) and Surge-Prioritization Error (predictor) ........................................................................................................87
Table 4-11. Summarization of Experimental Variables .................................................................88
Table 5-1. Summary on effect of learning .................................................. 100
Table 6-1. Set of contributions from the study ............................................. 113
ACKNOWLEDGEMENTS

I am thankful to many faculty-members, colleagues, family and friends for their support during the course of my doctoral program.

Firstly, I would like to thank my advisor Dr. Michael McNeese for his support and encouragement throughout my graduate program. I thank him for sharing his knowledge and time that has shaped my dissertation research as well as my career goals and professional development. I am also thankful to Dr. McNeese’s feedback on my dissertation drafts.

I would like to thank Dr. David Hall for his support and, time in providing detailed feedback on my dissertation.

I would like to thank Dr. McNeese and Dr. Hall for supporting me financially through research grants during my graduate program. It was a great opportunity working with Dr. Hall and Dr. McNeese on various projects and learning through various discussions about issues that pertain to conducting and managing research.

I would like to thank Dr. Luke Zhang for his encouragement and support through various phases of this dissertation. I thank him for sharing his knowledge and time that shaped my experiment, data analysis and discussion. His advice on career planning and professional development is highly appreciated.

I would like to thank Dr. Ling Rothrock for his feedback and pointers that shaped my dissertation. His pointers on chapters involving methodology and statistical analysis and suggestions about implications from this study are highly appreciated.

I would like to thank, MINDS group and NC2IF members Katherine Hamilton, Vince Mancuso, Nick Giacobe, Mike Tyworth, and Eric McMillan for their co-operation and support along multiple dimensions involving running the experiment, developing the simulation, working
out logistical issues and, refining various materials and instruments utilized in the experiment. This study could not have been possibly completed without their support and encouragement.

I would like to thank various faculty members and instructors at the College of Information Sciences and Technology for their generous support in providing the opportunity to recruit participants from their courses for extra credit. I would like to thank Dr. Edward Glantz for his co-operation and encouragement.

I appreciate the support from various members of the administration at the College of Information Sciences and Technology, including Sue Van Vactor, Rita Griffith, Wendy Fay, Sue Kelleher, Michelle Hill, Brandy Bower, Debbie Kozel and Betty Blair for their support and cooperation during my graduate program and Karen Brewster who helped with grammatical edits to this dissertation. I appreciate their professionalism.

Last but not the least, I am thankful to my family who have provided me their support throughout. I also thank various friends and colleagues who I met at Penn State for their co-operation and support. Their company positively affected my progress.

This work was partially supported by U.S. Army Research Office (ARO) MURI Grant “Computer Aided Human Centric Cyber Situation Awareness” W911-NF-09-1-0525.
Chapter 1
Introduction

Cyber-security analysts are burdened with large volumes of data in the form of alerts from intrusion-detection systems, firewalls, vulnerability scanners and other network monitoring tools. They monitor organizational networks for threats that may disrupt organizational activity. The increased sophistication and recurrence of cyber-attacks in recent years has made the task of protecting cyber infrastructure critical to many organizations that depend on Information Technology. Human Factors issues pertaining to processing large amounts of data, prioritizing tasks, and team-coordination add to the challenge faced by cyber-security analysts (D'Amico, Whitley, Tesone, O'Brien, & Roth, 2005).

Task-prioritization and effective multitasking are a challenge within other domains as well, including power-plant monitoring (Waller, Gupta, & Giambatista, 2004), aviation (Bishara & Funk, 2002; Schutte & Trujillo, 1996), and emergency-response (Wellens, 1993). The ability to multitask is directly related to a variety of factors such as display design, presence of technological aids, organizational goals, and internal factors such as the operator’s experience on the task and individual differences.

This study aims at understanding the effect of workload-previews on task-prioritization and task-performance. Previews are look-aheads that predict the future state of the environment. This may include workload, severity or other information that may be useful for advanced mental preparation or cognitive readiness. Previews should facilitate making necessary adaptations for changed conditions in the future. Previews may improve attention-allocation and task-performance however, task-performance may also suffer with its presence. Experiments have shown that predictors that provide long-term predictions such as weather forecasts (Andre, Heers,
& Cashion, 1995; Segal & Wickens, 1990) or traffic reports are likely to improve decision-making and task-prioritization.

However, previews in some studies, particularly short-term previews providing dynamic updates on a future state, have shown to be detrimental to performance for reasons that relate to the increased cognitive-load resulting from the need to integrate preview information in task-management in a short time-frame (Cummings & Mitchell, 2005; Parasuraman, 2005; Wickens, Pizzaro, & Bell, 1991). In a study on short-term predictions in multiple-UAV task-management, the provision of an aid that predicted workload-bottlenecks had a negative effect on task-performance as participants appeared to have re-planned targets more often than they should have. This possibly resulted from attempts to optimize schedules for visiting targets (Cummings & Mitchell, 2005).

This study explores the effect that short-term predictions of event severity and event volume would have on task-prioritization and task-performance in a cyber-security event-monitoring task. The short-term predictor is being explored within the context of a dual-task environment in which analysts may be overloaded with event related data. The study revolves around two Human Factors issues – why individuals fail to switch attention to more critical tasks, and what effect workload-previews have on task-prioritization and task-performance. The first issue relates to research in task-management and interruptions that have explored factors that influence the probability of not attending to a high priority task (Wickens & McCarley, 2008). Studies on cognitive-tunneling, change-blindness, task-management, task-prioritization and modalities of interruption have been extensively reported. The second issue concerns the effectiveness of previews or predictors on performance in time-pressured decision making tasks.

Significant research has focused on the reliability of previews. Predictors with a high false-positive or false-negative rate may be ineffective in facilitating task-performance. They may also be detrimental to task-performance (Metzger & Parasuraman, 2005). Information overload,
high false-alert rate, or high miss-rate may contribute to degraded performance. However, the focus of this study is not on reliability. The results of this experimental study would have implications to literature concerning task-prioritization, workload-preview design format, cybersecurity monitoring and decision-making. Workload-previews have shown to be detrimental to task-performance for a variety of reasons as has been reported in previous research (Cummings & Mitchell, 2005; Wickens, et al., 1991).

In this research, the experimental task consists of several discrete tasks within each sub-task (primary and secondary). Participants are required to respond to discrete cyber-security events within an organizational network. The primary task is the ‘default task’ as is instructed to participants. The secondary task, however, will take precedence over the primary task interaction, if severity of secondary task events exceed that of the primary task. This simulates realistic dual-task contexts where a secondary task may occasionally present analysts with unexpected and critical alerts that require urgent attention.

The predictor in the task provides information on future overall severity of sub-tasks that facilitate timely switching to the secondary task. The preview accompanies the information of current and past average threat-levels. If this display of information results in exceeding the task-performance achieved by only having current threat-level information, the study would indicate that task-prioritization performance is enhanced with the availability of early-warnings or previews. Furthermore, if this benefit is more pronounced within a higher workload condition, the study would indicate that previews have the ability to mitigate the deteriorating effect of workload on task-performance. On the other hand, if performance levels remain similar, or if previews lead to poorer performance, this would suggest that display designers should re-consider the use of short-term predictors. This is especially important as operators and analysts in several domains involving surveillance and threat monitoring are overloaded with decision-aids, visualizations and warning systems that may not get used.
Motivation: Cyber-Security Event Monitoring

Cyber-security analysts are responsible for protecting organizational computer-systems including software, hardware, and data from external threats. The responsibilities of cyber-security analysts differ based on the organization, its size, its policies and its technology infrastructure. Typically, cyber-security analysts monitor network-traffic to detect malicious activity. Computer Network Defense (CND) analysts, specifically, are provided with data from intrusion-detection systems (IDSs), firewalls, and other network monitoring systems. At a very high level, tasks in cyber-security analysis can be classified into reactive tasks and proactive tasks; Reactive tasks are related to sense-making from log-files or alerts from automated systems, or identifying patterns present within multiple cues that are associated to events that have already occurred; Proactive tasks involve the sense-making processes to predict a potential incident or attack (D’Amico & Whitley, 2008).

D’Amico and Whitley (2008) conducted a cognitive-task-analysis on 41 CND analysts in seven organizations, which involved semi-structured interviews, hypothetical scenario-construction, review of critical incidents and naturalistic observations. As IDSs initially generate large amounts of raw data, the authors indicate that this raw data undergoes a process of filtering and integration through various cycles. They describe the process of developing Situation Awareness as an iterative process involving raw data being filtered into network behavior of interest. That data is further filtered into suspicious activity that further leads to the identification of specific events that may be worthwhile reporting. The process is time-consuming and these stages can involve data collected over several months. Given the large number of sources that are used to detect threats and analyze data to locate interesting activity, analysts are required to work in teams and attend to multiple sources of threat-data. It is also likely that analysts would be required to maintain prescribed task-priorities where some types of threats, or some types of alert
information, should be given more importance than others. Contextual factors, such as a recently discovered vulnerability, exploit or the specific assets at risk would require analysts to re-assign priorities. It should be noted that cognitive-task analyses conducted by that of D’Amico and Whitley (2008) may be among several others, however, these reports are difficult to obtain given their classified and proprietary nature.

Cyber-security analysts may benefit with predictive information in addition to alerts generated by intrusion-detection systems and firewalls. Facilitating analysts with predictive information may provide them with the relative risks associated to various sub-nets within the organizational network. Such information should affect attention-allocation, improved decision-making, and thereby improved task-performance. Cyber-attacks are often a blend of various procedures and methods combined. These are better known as multi-stage attacks that happen in distinct stages of progress before the actual goal can be accomplished. Identified precursor events may be capable of predicting a blended attack or multistage attack. Precursor events for assessing future threat may include zero-day attacks, anomalies in network behavior (Wang, Cretu, & Stolfo, 2006; Wang & Stolfo, 2004) or cues indicative of repetitious and malicious behavior by cyber criminals.

There have been several attempts to develop systems that would provide analysts with early warnings on potential threats (Friedrichs, Levy, Huger, & Tomic, 2003; Liljenstam, Nicol, Berk, & Gray, 2003; Yang, Byers, Holsopple, Argauer, & Fava, 2008). An example is a proposed system to detect, early on, the spreading of worms across the globe based on a trend detection methodology (Zou, Gong, Towsley, & Gao, 2005). According to a study conducted by Narus Inc., security personnel have indicated the need for early warnings about potential threats and attackers, and more information leading to the prioritization of specific threats and attackers (Narus, 2011).
Stolfo (2004) addresses the issue of developing techniques to detect malicious reconnaissance activity for being able to predict potential cyber-attacks. Wormonitor is a project addressing several techniques to detect, report and defend against cyber-attacks. The project addresses the ability to provide early warnings for worms and precursors to zero-day attacks, and, the ability for network anomaly detection systems to discover new malicious attacks and automatically make these systems generate zero-day attack signatures. Various issues pertaining to the ability to generate early warnings of cyber-attacks is addressed in the project (Stolfo, 2004).

Gu, Sharif, et al (2004) report a behavior based model to detect worms and generate early warnings to prevent their spreading. They state that previous approaches for detecting worms rely on worm’s rapid scans for vulnerable hosts. These scans usually cover numerous inactive IP addresses. An unusual surge in the number of inactive IP address scans would be precursors to the spreading of worms; Some approaches would identify infection like behavior in vulnerable hosts. The authors propose a novel two-phase worm detection algorithm named Destination-Source Correlation (DSC) to detect worms. The DSC algorithm focuses on worm scanning behavior as well as the behavior of infected hosts to detect worms and provide early warnings about their spread. The study tests the effectiveness of early warnings generated from local host information, and the authors report that using their approach, early warnings for worms can be generated when only .19% of vulnerable hosts get infected by the worm (Gu, et al., 2004).

Future-state predictions on potential threat and Workload may be useful to cyber-analysts however, its ability to improve task-performance requires investigation from a Human Factors standpoint. Workload-previews have not necessarily improved performance in other tasks or domains (Cummings & Mitchell, 2005; Parasuraman, 2005; Wickens, et al., 1991). This study is particularly investigating the effect that previews have on performance from the perspective of attention management.
Research Framework

The investigation on the effect that predictors have on workload and threat-levels in cyber-security monitoring can be explored from many perspectives, levels of analysis and disciplinary frameworks. The level of analysis in terms of effects on human performance considered here is the individual level. However, studies exploring behavior within teams or organizations may benefit from this research. The effect of the technology on human performance can also be studied at various levels along the scale of time. Issues such as automation reliance, fatigue, work-shifts, development of expertise and other long-term processes, although relevant to cyber-security, are not considered here. The benefit of the workload-preview is being studied at the scale of a few minutes. The literature review in this dissertation has addressed previous research in the Human Factors, Psychology and Human-Computer Interaction (HCI) relevant to this dissertation. The problem being studied relates to task-management, task-prioritization, attention-allocation, workload and interruptions. Figure 1-1 provides an overview of the framework for this study. Cyber-security event-monitoring is a complex task that can be seen from the perspective of Macrocognition (Klein, et al., 2003; Letsky, 2008). The Macrocognition framework identifies several macrocognitive functions and supporting processes that operate in a naturalistic environment. This dissertation mainly focuses on the supporting process of Attention Management. The functions of adaptation/re-planning also relate to the re-prioritization of tasks – an issue of interest in the study. While cyber event monitoring can also be seen from the perspective of MetaCognition (Carruthers, 2009), a research area that addresses cognitive processes involving the reflection and interpretation about an individual’s own thought processes, this dissertation does not consider this perspective.
The Effect of Workload Previews on Performance in Cyber-Security Event Monitoring

Task-Management, Interruptions, Attention Allocation, Workload,

Hypotheses

Experimentation/Analysis

Implications to Theory

Implications to Workload-Preview Format

Implications to Cyber-Security

Figure 1-1. An overview of the framework followed in this experimental study

Research on external-interruptions in various domains is related to the problem addressed here. This area has mostly looked at how an interrupting-task disrupts cognitive processes involved in an ongoing-task. A number of factors have been explored such as effects of varying mental-workload (Iqbal & Bailey, 2005), the availability of perceptual-cues at task suspension and task resumption (E. Altmann & Trafton, 2002; Hodgetts & Jones, 2006), various characteristics of the interrupting-task such as similarity (Czerwinski, Chrisman, & Schumacher, 1991), modality (Ho, Nikolic, & Sarter, 2001; Ho, Nikolic, Waters, & Sarter, 2004; Latorella, 1998; Smith, Clegg, Heggestad, & Hopp-Levine, 2009), duration of the interruption (Monk, Trafton, & Boehm-Davis, 2008), among other factors (Gillie & Broadbent, 1989; Monk, Boehm-Davis, & Trafton, 2002). The problem explored in this study has similarities to this area however,
it more closely relates to task-management in which the distinction between the interrupting-task and ongoing-task is blurred, and issues such as task-prioritization strategy become important (Wickens & McCarley, 2008). Rather than an interrupting-task and an ongoing-task, the study considers a dual-task environment where both tasks are similar in nature, and they only differ in their respective sub-task priorities. Moreover, the secondary task is not available in the form of a forced interruption.

Task-Switching in Psychology is a related area and the term generally refers to a paradigm in experimental-psychology that explores mental-processes of cognitive control underlying switching between tasks. Task-Switching research involves the analysis of response-time differences that are within the order of a few 100 ms in stimulus-response tasks such as distinguishing whether a digit is even or odd. While these studies provide interesting results relevant to theories in cognitive control, the results from these studies are often not generalizable to real-world tasks. An in-depth overview of task-switching is available in a review of this research area (Monsell, 2003).

Significant research has explored prospective memory issues in mission critical tasks (Dismukes, 2001; Dismukes & Nowinski, 2007). This area explores failures to recollect intentions from memory after being presented by an interruption. The issue being explored in this study does not take into consideration failures or errors in relation to task-resumption or formation of intentions in memory. The effect of workload-previews on task-prioritization and task-performance is studied from the perspective of task-management, interruptions, attention-allocation and workload. The hypotheses being tested through experimentation are based on previous research on task-prioritization and effects of workload-previews mainly reported in the Human Factors, Psychology and HCI.
Chapter 2

Literature-Review on Attention, Task-Management and Workload-Previews

A variety of cognitive issues can be explored in cyber-security event monitoring. This study addresses human ability to monitor concurrent tasks that change in relative severity with time. Monitoring organizational infrastructure involves addressing several locations and assets. It is essential for analysts to classify these locations and assets into priority-levels. Analysts are faced with the challenge of being able to sufficiently allocate mental resources across a range of tasks that differ in priority level. This study specifically assesses the effect of a workload-preview on the analyst’s ability to selectively attend to tasks and take notice of occasional changes in relative severity. This effect is examined at varying levels of task-load as human behavior can change with variation in task-load. The literature review covers this from the perspective of attention theory, mental-workload and Situation Awareness. Experimental research on interruptions, task management, and workload-previews is discussed in separate sections. This is followed by a section providing an integration of the literature that presents the Research Question and Hypotheses.

Attention Theory and Workload

Attention as a subject in Psychology and Human Factors focuses on the mental processes that underlie the application of cognitive resources to a variety of tasks carried out on a day-to-day basis. This section provides an overview of theoretical frameworks on Attention that have been used in a variety of contexts such as time-pressured decision-making, driving, aviation, etc.
Attention is one of the most well studied areas within Psychology. Attention as a term, represents a number of related cognitive processes. There have been various attempts at defining Attention. According to William James, “Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German.” (William, Frederick, Bowers, & Ignas, 1890).

Wickens (2007) explains attention with the help of two metaphors – attention as a spotlight and attention as a limited-resource. Attentional processes underlie complex human behavior demonstrated on a regular basis. Driving while speaking and watching a television show while cooking a meal are a couple of examples. The success of these complex behaviors underlies what is also responsible for cases of failure. Losing track of time with a known time-limit on a parked car; or failing to take notice of a braking vehicle on a street while driving and texting are some examples of attentional failures.

The variety of cognitive failures that can be attributed to attention corresponds to the varieties of attention stated by Wickens (2007) as a classification of various attentional functions. They are divided-attention, focused-attention, sustained-attention, and selective-attention. The two metaphors that relate to attention represent different mental processes. The SEEV (Salience, Effort, Expectancy and Value) model addresses selective-attention or the ‘Attention as a Spotlight’ function of attention.

During 1991 in Los Angeles, an air-traffic-controller positioned a SkyWest Metroliner on a runway and switched attention to other airplanes. Without repositioning the SkyWest, the same controller cleared a USAir Boeing 737 to land on that runway that resulted in the collision of the two airplanes fatally injuring several passengers and crew members. There are several factors that
are believed to have contributed to not attending to the SkyWest that was supposed to be repositioned. Wickens and McCarley (2008) have described it as a failure of selective-attention. The inattention has been associated to a number of contributing factors including an unpredictable interruption that occurred regarding communication with another aircraft, poor visibility (from the control tower) resulting from glare (after sunset), and the slow arrival of a flight-strip of a Wings West aircraft to the controller (NTSB, 1991).

The SEEV Model

The SEEV model is a model of selective-attention (C. D. Wickens & J. S. McCarley, 2008). The model incorporates top-down and bottom-up factors that influence selective attention towards specific Areas-of-Interest (AOI) in a visual environment. Its mathematical model estimates transition probabilities from one AOI to another within the visual environment. It predicts visual attention-allocation based on four factors – salience, effort, expectancy and value. Salience and effort are bottom-up factors – the eyes tend to move towards salient events, however, the effort required for that movement may negate the salience factor. The distance required to complete the visual-scan constitutes effort among others. Visual scans are also likely to occur on locations where the probability of finding information may be high (expectancy). This likelihood would be even higher when such information has high value. Both expectancy and value are top-down factors and they are influenced by the knowledge. Expectancy of sampling a particular channel is driven by event-frequency along that channel and contextual cueing.

The SEEV model can be applied in training for appropriate scan patterns, display design considerations, safety prediction by predicting attentional neglect, and design of highly reliable automation systems. However, the SEEV model has not been successful in predicting behavior relating to non-optimal allocation of attention. Research has not yet explained why individuals
cannot switch from a “compelling task” to a critical task of high-priority or why individuals get
distracted away from a high-priority task to other compelling tasks. According to Wickens (2007)
the challenge remains in defining the factors that govern task-management in the real-world.
Also, the SEEV model does not address the issue of mental-workload and the influence that
mental-workload has on selective-attention.

Inattentional Blindness and Change Blindness are two among other types of failures of
selective attention. Inattentional Blindness is the looked-but-failed-to-see effect where an
important event was prominently visible, but the operator still failed to detect it. This looked-but-
failed-to-see effect is a cause for a number of traffic accidents (Wickens & McCarley, 2008).
Patterson in a pediatric nursing study provides an example of inattentional blindness where an
inexperienced nurse fails to notice visual cues on an infant. The visual cues indicating the need
for urgent care was later noticed by a more experienced nurse (Patterson, Ebright, & Saleem,
2011). Change blindness is a failure to detect that an object in the environment is different from
what it was. An example is where an air-traffic-controller fails to notice altitude or speed changes
on an aircraft represented on the display while attention was being focused on a different area of
the display (Wickens & McCarley, 2008). Understanding the influences of various factors on
accidents relating to attention could inform display-design, automation-design and operator-
training. The experiment in this study requires subjects to notice changes in severities of sub-
tasks being monitored in a timely manner. The inability to notice these changes in time would be
an instance of change blindness.

The Multiple-Resource Theory

The Multiple-resource theory addresses the ‘Attention as a Resource’ view of attention,
and serves as a basis to issues in divided-attention. It explains why some time-shared task-pairs
result in greater interference than others. For instance, performing a spatial processing task would interfere with driving performance more than attending to a phone call. This would vary with the level of expertise in driving, familiarity of context of the phone conversation as well as that of the spatial processing task. The Multiple-resource theory is also represented by a model that predicts multitask interference between tasks being concurrently performed (Wickens, 2008). The validation of this model has been reported in multiple articles (Horrey & Wickens, 2003; Sarno & Wickens, 1995; Wickens, Dixon, & Ambinder, 2006). The model has successfully predicted data produced with human experiments on high-fidelity driving simulations involving various driving and dual-task conditions (Horrey & Wickens, 2004). It particularly predicted hazard-response (98% variance accounted) and in-vehicle driving performance (92% variance accounted).

Wickens however states that the model could not predict lane-keeping performance, as the model used in the analysis did not accommodate the separate resources of focal and ambient vision; and resource-allocation policy that represents a preference for one task over another.

The view that human performance is bounded by a resource-pool of mental-effort (Kahneman, 1973) has inspired the multiple-resource theory. The theory relates to the concept of mental-workload and breakdowns in dual-task performance under high workload. The extent to which time-shared task-pairs interfere depends on the number of mental resources they share. Some tasks are more mentally demanding than others. Tasks that are automatic, such as walking, would be less mentally demanding than complex tasks that may have not been learned. Therefore, the extent of interference in time shared tasks would also depend on the mental-effort demanded by each task.

The Multiple-resource theory posits that different stages in information-processing are supported by different resources. The perceptual and cognitive stages are independent of the response-selection and execution stages as shown in Figure 2-1. This implies that mental functions that mainly underlie the acquisition and maintenance of situation awareness have less of
an overlap with the mental functions involving response selection. Therefore, acquiring and maintaining SA and selecting responses would show relatively less interference for resources. On the other hand, decision-making would interfere with SA relatively more.

Response-selection and execution vs. perceptual-cognitive processing relates to the stages dimension within the resource model. Wickens also includes modalities, and codes of processing dimensions.

The Multiple-resource theory has been supported by neurophysiological data (Wickens, 2008). Although, the theory has been able to predict multiple-task performance for a number of experiments, the motivation underlying the theory has been towards predicting the utility of one type of interface over another in terms of dual-task processing. Wickens (2008) also states that predictions are accurate for (dual) task-demand at or beyond the red-line of workload, rather than situations where demand remains at ‘residual capacity’.

![A Model of Human-Information Processing Stages (Wickens & Hollands, 2000)](image)

Figure 2-1. A Model of Human-Information Processing Stages (Wickens & Hollands, 2000)

In Figure 2-1, the blocks with dashed borderlines are separated into two colors – This separation mainly indicates that these functions are supported by separate resources
corresponding to the information-processing stages dimension in the multiple-resource theory (Wickens & Hollands, 2000).

The multiple-resource theory is yet to address several challenges including accommodation of tactile input as a modality and importantly what drives resource-allocation policy. In other words, the theory does not address failures of attention allocation as seen in auditory preemption or the use of cell phones while driving (Wickens, 2008).

**The Situation Awareness Theory**

Situation awareness and the processes that underlie acquiring it are related to workload, attention, expectancies and allocation policy. Good situation awareness is generally associated with high performance and effective task-management. The situation awareness model provided by Endsley and Garland (2000) can be considered to be a theory of Attention. The task of event-monitoring in cyber-security involves attending to alerts generated from various systems and accordingly detecting and identifying threats. It is an interplay of different mental processes including perception, attention and memory. The situation awareness model provides an overview of various factors and elements involved in this complex interplay underlying successful event-monitoring in the cyber-security context (Endsley & Garland, 2000).

The definition of situation-awareness has been debated among researchers, however, a well-accepted definition is provided by Endsley M.: It is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (Endsley, 1988). Situation awareness, as the definition suggests, consists of three levels – perception, comprehension and projection.

According to Endsley’s definition, situation awareness refers to the state of knowledge or the outcome of several cognitive processes and not the processes themselves; the processes
working towards acquiring the state of knowledge is referred to as situation assessment. These are
interrelated as attentional processes underlying situation assessment are directed by the
knowledge of the environment and vice-versa.

Endsley states that a good prioritization strategy within a dynamically changing
environment is a challenge to acquiring situation awareness. The spotlight of attention, or
selective-attention, is influenced by bottom-up and top-down factors. Task goals guide top-down
attention where elements in the environment get filtered and information relevant to goals is
processed at a perceptual level and may later be processed at the cognitive level. However,
bottom-up factors can disrupt the influence of top-down processes with the appearance of salient
cues in the environment. Without bottom-up drivers of attention, cyber-security analysts would
fail to notice unexpected shifts in priority, or a driver looking at a navigational-device may fail to
notice a braking vehicle ahead on the highway. There is no doubt that switching between the two
modes of attention - top-down and bottom-up, is necessary for maintaining situation awareness.

The three levels of SA in Endsley’s definition based on the above discussion on how SA relates
to attention, mental-workload, and memory are provided below.

**Level 1 SA:** It refers to the perception of elements or cues in the environment that is relevant to
the task. Elements or events in the environment that have entered the analysts perceptual
registers, working memory, or long-term memory may not necessarily be utilized unless the
analyst can integrate these elements together towards meeting the objective(s) of the task.

**Level 2 SA:** SA is not merely the perception of the elements or cues. The comprehension of their
meaning and how they are relevant to current goals of the task is an integral part of SA. Top-
down attention directs selective-attention towards areas that are relevant to current goals, and
following perception data gets integrated for the comprehension of the state of the environment.
This integration requires cognitive resources such as working-memory, long-term memory or the
problem-state resource (Salvucci & Taatgen, 2011). Under conditions of high mental-workload,
these cognitive resources become depleted leading to a break-down of the processes necessary for appropriate comprehension. Bottom-up attention may be sufficient to develop level-1 SA, but comprehension may require top-down attention processes, unless the difficulty of the task matches with the analysts’ level of expertise such that all comprehension required from the task can be acquired through pattern-matching. Direct pathways exist between sensory registers and long-term memory to facilitate such pattern-recognition (Cowan, 1988; Endsley & Garland, 2000) or event recognition-primed decision-making (Klein, 1993).

**Level 3 SA:** The ability to project or forecast future events is an indicator of expertise within the domain in question. A superior level of comprehension about the state of the environment entails the ability to project events that could unfold in the near future. This is also related to developing expectancies that are necessary for goal-directed selective attention.

Lower levels of SA are necessary for the development of higher level SA. Expectancies or the ability to project future state is a result of good comprehension. This study investigates the implication of providing a workload-preview or a predictor on task-performance. Situation awareness plays a mediating role between technology and performance. Technology providing information about the future state includes cues that would lead to better comprehension. However, if the technology has a tendency to throw operators out-of-the-loop (Endsley & Kiris, 1995) level-2 SA or comprehension may actually suffer. In this study, the latter is not an issue of concern as the preview itself is not enough for successful task-performance. Participants will have to process information provided about events to identify their categories regardless of whether they have a preview or not.

The above discussion suggests that if task-management within this experimental task depends on perception (i.e. less on comprehension or information integration). Irrespective of whether a workload preview is provided, cognitive and temporal resource-allocation would be near-optimal under high workload, as perception should not suffer under high workload.
Additionally, the provision of the workload-preview may not result in any benefits either. The relationship between workload and task-prioritization has been studied and reported in the Psychology and Human Factors literature within many contexts, and task-prioritization does suffer under high workload. Moreover, top-down drivers of selective-attention suffer under high-workload. The sub-section on workload effects on attentional processes addresses this in more detail. The next sub-section provides an overview on the concept of mental-workload that captures the relationship between task-load and mental-effort.

**Mental-Workload**

The demand to manage multiple tasks and perform them based on prescribed priority order influences task-performance, although this may vary based on the type of task and the extent to which it utilizes cognitive or attentional resources. Some tasks are more “disruptable” than other tasks for an individual. Cyber-analysts are required to make sense out of a large number of reports that sensors generate, and there are situations where cyber-analysts experience high levels of mental-workload. In a task-simulation, task-load may be increased, however, this may not necessarily reflect in increases in cognitive-effort. Mental-workload captures the relationship between task-load and the mental-resources of the operator (Wickens & Dixon, 2007). As task-prioritization varies with mental-workload, it would be beneficial to measure mental-workload.

A non-operational definition of mental-workload is provided by O’Donnell and Eggemeier as “that portion of the operators limited capacity actually required to perform a particular task” (O’Donnell & Eggemeier, 1986). This definition relates to the notion of “limited capacity” and is similar to a capacity theory of attention. Wickens and Dixon define workload for a human operator as the “load imposed on limited resources of the unaided operator”. They state
that the source of this load can be from the single task difficulty or from concurrent task load (Wickens & Dixon, 2007). Reid and Nygren (1988), describe workload to be a multi-dimensional construct that can be explained by time-load, mental-effort load, and psychological stress load. They conducted a literature-review, and noted 20 definitions of mental-workload, and also identified three factors as most commonly used by authors; (1) time load (2) mental effort load, and (3) psychological stress. Time load is referred to as the time-available and task-overlap among tasks to be accomplished. Mental-effort load is determined by the information processing performed. It involves information processing such as calculations, short term memory usage, and decision-making. Psychological stress is described as any source of anxiety, confusion or frustration (Reid & Nygren, 1988). In Experimental Psychology, mental load is defined by Moray as “the rate at which information is processed by the human operator, and basically the rate at which decisions are made and the difficulty of making the decisions” (Moray, 1979). It has been referred to as - where ‘human operator workload’ is divided into three attributes that functionally relate to workload – input-load, operator effort, and performance of work result (Jahns, 1973).

Sheridan and Stassen state that operator workload has been defined to mean different things such as the physical task assigned; the demands of the task in terms of criteria, decision-risks, attention and temporal requirements; information processed by the operator; energy spent by the operator; emotional stress; and system performance. They state that system performance is not workload. Additionally, they also state that information-processing and emotional-stress, together, is referred to as “mental load” (Sheridan & Stassen, 1979). Cognitive-Load Theory in the Educational Psychology area is based on the idea that a major factor of the effectiveness of instructional formats is the memory load as it results from the instructional format (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Paas et al., note that Cognitive-Load is a multi-dimensional construct that refers to the “load that performing a particular task imposes on the
learner’s cognitive system”. They also state that it consists of aspects of mental load, mental-effort, and Performance. Mental load refers to the “cognitive load that originates from the interaction between task and subject characteristics”. Mental-effort refers to the allocated cognitive capacity to meet the demands of the task, while Performance is the learners’ achievements or the time to perform the task (Paas, et al., 2003).

The concept should tie the relevant dimensions together and exclude ones that are not essential. Many authors agree on mental-workload as a demand placed on mental resources of the operator. The temporal component also impacts the demand placed on mental resources for a given time-period. Even though demand placed on limited resources emerges from the intrinsic characteristics of the individual and the task, mental workload or cognitive load has been argued to be what is experienced by the operator at the time the task was performed. There is a general disagreement on whether performance is a component of mental-workload. High mental-effort may lead to high performance, however, the inverse is not true. Therefore, performance may be discounted. While psychological-stress may be correlated to mental-workload, it is not considered by a number of researchers.

Mental-workload can be defined here as a concept that refers to mental-effort exerted by the individual in a given amount of time in order to meet a given demand. Mental-effort also relates to energy spent by the operator, since applying mental-effort requires neuronal activations in corresponding areas that employ the necessary mental-resources. Attention as a concept relates to mental-workload or mental-effort however, it basically encompasses the mental-processes and cognitive-resources underlying the production of goal-directed behavior. In short, mental-effort and the energy spent is the result of various attentional-processes that work towards satisfying a goal.

There have been several attempts at operationally defining mental-workload. Schnotz and Kurschner state that cognitive load is composed of intrinsic load, extraneous load and
germane load. Intrinsic load is the component of cognitive load that results from natural complexity of the task, which involves, element interactivity or the number of elements that need to be placed in the working memory to process the task. Extraneous load is cognitive load that results from the instructional design and is unnecessary to the process of the actual learning. Germane cognitive load is caused by schema construction and schema automation that can result in lower intrinsic load and lower extraneous load (Schnotz & Kürschner, 2007). It has been stated that performance in a secondary task can be a measure of workload in a primary task, within a range that lies below operator overload (Wickens, 2002). Similarly, Paas et al., have suggested that secondary task performance is a good indicator of cognitive-load experienced on the primary task (Paas, et al., 2003). Mental-Workload can also be operationally defined as a variable which is equal to the time required to perform tasks divided by the time available to perform tasks (Wickens, 2002).

Bunce, Izzetoglu et al (2011) conducted a study to examine the relationship between hemodynamic response in the prefrontal cortex, mental-workload and expertise in a task involving multitasking and vigilance. They conducted an experiment with 8 participants. They were placed in a command-and-control task at various task-load levels. At low and moderate levels of task-load experts showed lower levels of oxygenation in comparison to novices, whereas, at high task load experts showed higher oxygenation levels. The authors suggest that novices disengaged from the task at high task-load and that oxygenation levels as detected with functional near-infrared spectroscopy were associated with applied mental-effort (Bunce, et al., 2011).

Mental-workload may also be measured subjectively. If an operator reports that he / she is overloaded or under stress, then, it should be concluded that he / she is overloaded, “regardless of what other indices might lead you to conclude” (Reid & Nygren, 1988).
There are several proposed measures of mental-workload reported in the literature. Subjective measures where participants need to respond to items in a questionnaire would have good *construct validity*. Paas et al. (2003) have stated that subjective measure for the construct would have good discriminate, construct and convergent validity. Secondary task-performance (Wickens, 2002) measures may have good face validity because it is easier to focus on a secondary task if the primary task does not overload the operator. However, secondary task-performance may be measuring distraction or the inability of maintaining an optimal prioritization strategy thereby becoming a source of noise to mental-workload. The ratio of time required to perform the task to the time available (Wickens, 2002) appears to capture the time-pressure component of task-load. However, it is difficult to determine optimal task time required to complete a task.

Webb et al (2010) proposed a measure of workload assessment based on a subjective survey in a study with Black Hawk pilots and instructors as eligible participants. Their assessment technique involves providing the survey on a retrospective basis - unlike other workload measurement instruments, they provide the survey after performing all tasks. The survey consists of items that require the participant to assess cognitive, visual, verbal, aural and physical demand on a 0 to 4 scale. The authors state that although the measure should be tested for correlations with validated workload measures (physiological and subjective), their measure was sensitive to various tasks and workload dimensions. A limitation of this retrospective approach according to the authors is that it is lengthy and individuals may have memory limitations on being able to report subjective workload retrospectively and accurately (Webb, Gaydos, Estrada, & Milam, 2010).

A measure of mental workload should reflect the dimensions of mental-effort subjectively experienced by the operator over a given period of time. For this study, a subjective questionnaire was judged to be a good instrument to measure mental-workload due to its discriminant, construct, and convergent validity (Paas, et al., 2003). Examples of such
instruments previously designed are the NASA Task-Load Index or NASA-TLX (Hart & Staveland, 1988) the Subjective Workload Assessment Technique or SWAT (Reid & Nygren, 1988) and the Workload Profile (Rubio, Díaz, Martín, & Puente, 2004). Rubio et al. (2004) state that the three measures have high convergent validity. Modifications can be made to these measures where questions that do not tap into the concept can be removed, such as the subject’s perception of task-performance.

**Workload Effects on Attentional Processes**

An operator conducting a task requiring the application of mental-effort reaching resource-capacity limits will demonstrate break downs within attentional processes. These break downs in attentional processes are related to issues in task-management.

High mental-workload is associated to selective attention being mainly influenced by bottom-up factors. Elements in a display, associated with high salience, or, that may be too cluttered, would often influence bottom-up attention thereby interrupting an ongoing task. Task-prioritization can be considered to be a separate sub-task or secondary task. It involves updating ones’ awareness about the status of individual sub-tasks in order to successfully allocate attention to the task with highest current priority. Secondary task performance for many tasks is considered to be a measure of cognitive-load on the primary task. The secondary task of task-prioritization would be negatively affected under high task-load, although the presence of affordances on displays that facilitate task-prioritization may moderate the effect of task-load on task-prioritization.

Demanet et al. (2010) examined the relative contributions of bottom-up priming factors and top-down control factors in task-selection. In experiments in this study, working-memory load was induced to reduce the effect of top-down attention (Logan, 2007). The researchers
manipulated bottom-up factors such as stimulus-repetitions, irrelevant-information and stimulus-task associations. These bottom-up factors introduced in the task influenced task-selection. According to the authors, top-down attention could only counteract the effect of stimulus-repetitions. The experiment showed that under memory load participants repeated tasks more often. The authors suggest that top-down attention is necessary to avoid the tendency to repeat tasks. This result indicates that the tendency for task-repetition increases with mental-workload which in this study was induced by working-memory load. The tendency to repeat tasks is similar to attention-tunneling. While this study indicates that attention-tunneling can be a result of high mental-workload, it suggests that this attention-tunneling was a result of increased bottom-up influences or reduced top-down control. They suggest that bottom-up attention drives task selection under conditions of high load (Demanet, Verbruggen, Lefooghe, & Vandierendonck, 2010).

Previous psychological studies examining the effect of different types of load on the ability to avoid the processing of visual-distracters suggests that under high-perceptual load, distracter processing would be reduced, however, load placed on cognitive control processes would lead to increased distracter processing indicating that the inhibition of distracters taps into resources of limited capacity (Lavie, 2005).

There have been similar studies on the influence of anxiety on goal-directed attention. Goal-directed attentional control is responsible for carrying out actions according to a plan stored in long-term memory. Anxiety can disrupt this process and cause the operator to perform sub-optimally. Studies have shown that anxiety can cause attention to be tunnelled. It may even be associated to the effect of being more vulnerable to distraction. Attentional tunneling is the effect where an operator’s attention is selectively focused on a particular channel of information for a prolonged period of time with other information being left unattended leading to negative consequences (Hudlicka & McNeese, 2002; Wickens & Alexander, 2009). Attentional tunneling
can also be seen as the inability to switch onto a secondary task from a primary task. There are various factors that cause attentional tunneling, and anxiety and stress have been cited as causes (Easterbrook, 1959). Attentional tunneling can cause operators to miss important cues leading to poor situation awareness and incorrect decisions or even fatal accidents in the context of car-driving where information about nearby traffic may be left unattended.

The studies above indicate that mental-workload is associated to reduced influence of top-down attention and that task-selection is mainly influenced by bottom-up factors under high mental-workload.

**Interruptions**

Significant work has been done in the Human Computer Interaction literature on interruptions and how factors such as workload, timing of interruption, interruption-lag, presence of a warning prior to the interruption, alert-modality and type of interrupting-task affect the ability to switch between an ongoing-task and an interrupting-task and vice-versa. This dissertation does have some similarities to studies investigating interruptions. As it explores the effect of a workload-preview on the ability to notice changes in relative priorities between sub-tasks, a factor of particular interest in this dissertation is the effect that a pre-interruption warning would have on the ability to switch to an interrupting task and resume from it. The effect of varying interruption-lag, the time-duration between the interruption-warning and the interrupting task, may also be useful to consider.

Altmann and Trafton (2004) conducted a study to investigate whether perceptual cues related to the task were associated to preparatory processes prior to being interrupted. The authors state that previous research has indicated the effect of interruption lags on preparatory processes. They conducted an experiment in which two variables are manipulated – interruption lag (2, 4, 6,
8 seconds) and the availability of cues on the primary task display during the interruption lag. The dependent variable was resumption lag. Resumption lag was calculated by considering the time-difference between the re-appearance of the primary task, and the first action (mouse click). The results of the experiment revealed that the availability of perceptual-cues facilitated task-performance at resumption, at least for the cases where resumption lags were longer. The authors suggest that preparatory processes may benefit from longer interruption-lags, which was 6 – 8 seconds in this particular experimental-task (Altmann & Trafton, 2004).

Trafton, Altmann et al. (2003) conducted a study to investigate the effect of the presence of an interruption warning on resumption lag. The objective of their study was to explore whether the cognitive system can take advantage of an interruption lag to reduce the resumption lag. They explored this from the theoretical perspective of retrieving information stored in memory at the time of the interruption. A failure to do so would involve the lengthy reconstruction of the goals and cognitive representations at resumption. Participants were placed in two conditions – a warning condition where an alert was provided 8 sec prior to the interruption, and an immediate condition where interruptions were abrupt. According to verbal reports, the participants in the warning condition prepared more than participants in the immediate condition. The participants in the warning condition also resumed the primary task more quickly. However, participants in the immediate condition improved in response times post interruption with practice. The authors suggest that the benefit of the interruption lag is that it provides a sufficient temporal window for prospective goal encoding, and retrospective rehearsal. The authors state that participants used the interruption lag towards better resumption where they focused on information available from the primary task or on encoding goals on what to do next. The warning impacted both primary task performance and the disruption-score. The authors state that their study has several implications. They suggest that interruptions should be accompanied with warnings and that even a 2 sec alert may be useful for preparatory processes for improved task resumption. Their pilot study included
the display of a time countdown after the warning. The authors state that the countdown may have interfered with preparatory processes, and it was thereby removed from the actual experiment. The authors also suggest that training operators for preparing to resume may be beneficial for post-interruption performance (Trafton, Altmann, Brock, & Mintz, 2003).

Iqbal and Bailey (2005) examined the use of workload-aligned models to identify opportune times for interrupting a task. Based on previous work (S. Iqbal, Zheng, & Bailey, 2004; S. T. Iqbal, Adamczyk, Zheng, & Bailey, 2005) the authors identified points in the task that cause the highest and lowest mental-workload measured with pupillary responses. These were classified as Best and Worst times to interrupt the primary task. In the experiment, the primary task was interrupted at Best, Worst and random times during the task. They measured resumption lag, annoyance and respect for the interrupting task. The authors reported that the predicted Best points in time were consistently associated to shorter resumption lags, less annoyance and more respect for the interrupting task. They state that small differences in the time when interruptions are presented can result in large mitigations in disruption caused by the interruption. An implication is that mental-workload can be an important factor that determines the extent of disruption caused by the interrupting task (S. Iqbal & B. Bailey, 2005). Similar studies on interruptions have shown that individuals have a tendency to delay interruptions to a low workload point in the primary task (Czerwinski, Cutrell, & Horvitz, 2000; Iqbal & Horvitz, 2007; Salvucci & Bogunovich, 2010; Wiberg & Whittaker, 2005).

Colcombe and Wickens (2006) conducted a study consisting of 4 experiments to examine the effectiveness of a conflict alerting system on performance in a dual-task context. The authors were interested in the effect of varying alert modality, the threshold of the alert and the level of resolution of the alert (binary 2-stage vs. likelihood 3-stage). In the first two experiments, the primary task of tracking, which was characterized by the authors as unstable, did not differ in the amount of engagement with variation of difficulty of tracking. Engagement was measured by
response-times to the interrupting alert. In experiment 3, the authors support their hypothesis that interruptions to an ongoing task towards the beginning of the ongoing task would be faster, in comparison to interruptions later in the ongoing task. This was mainly because of reluctance to leave an ongoing task on which considerable investment of mental processing has been made. The authors predicted that an increase in sensitivity of the alert threshold (i.e., increased false-alarm rate, reduced miss-rate) would result in delayed response times and poorer accuracy within the conflict detection interrupting task. While an increase in response-times was observed, a decrease in accuracy was not observed as delayed responses made it easier to discriminate between safe and unsafe trajectories. On the basis of Multiple-Resource theory (MRT), the authors hypothesized that cross-modal combinations (auditory-visual) and cross-code (verbal-spatial) combinations would be more compatible than within modal and within code combinations in terms of performance. According to the authors, results from experiment 1, 2 and 3 supported the MRT hypothesis. Based on the results, the authors provide recommendations for designing alerting systems (Colcombe & Wickens, 2006).

Mark, Gudith et al. (2008) conducted a study to examine the effect of context of interruption on performance in the primary task. The context was a within-subjects factor that had three levels including a baseline (no interruption), same context interruption and different context interruption. The primary task involved playing the role of a human resource manager that involves answering emails within the context of the role. The interruptions were questions presented by a co-worker. The context of the questions differed with the within-subjects condition of the experiment. Dependent variables included time to complete the emailing task, subjective-workload (NASA-TLX) and the politeness of email messages. The authors reported that the different contexts of the interruption (excluding baseline) did not effect the time to complete the primary task. Conditions where the task was interrupted resulted in faster completion times on the primary task. It is reported that stress, frustration, time-pressure, and amount of effort were rated
to be significantly higher in the interruption conditions in comparison to baseline. The authors state that individuals tend to compensate for interruptions by working faster by applying more effort which results in increased stress and frustration (Mark, Gudith, & Klocke, 2008).

The studies on interruptions discussed in this section address the relationship between interruptions and workload and the effect of warnings prior to interruptions. These issues would have implications on the study conducted here, although there are significant differences in context and the nature of multitasking behavior being investigated. Although the workload-preview discussed in a later section shares some similarity with warnings prior to interruptions, the workload-preview differs from warnings in many ways. Moreover, the secondary task considered in this study is not a forced interruption.

Breakdowns in selective-attention that have been responsible for many accidents in aviation happen under conditions of high workload in tasks where it is impossible to concurrently process several channels of information. This is viewed as sequential multitasking behavior where the distinction between the interrupting-task (IT) and ongoing-task (OT) is blurred unlike in the OT-IT-OT framework in interruptions research (Wickens & McCarley, 2008). The task being studied here significantly differs from the ones used in interruption related studies in HCI, as the primary and secondary tasks in this case are almost identical but, they represent different parts of the organizational network. The next section describes Human Factors issues in task-management and task-prioritization.

**Task-Management, Workload and Cognitive-Tunneling**

Research in the area of task-management in Human Factors more closely represents the issue in event-monitoring being studied as discussed previously. The distinction between an interrupting-task and an ongoing-task is blurred (Wickens & McCarley, 2008) in this study as
primary and secondary tasks are nearly identical, and they only differ in relative priority. This sub-section addressed research in task-management that specifically addresses task-prioritization in the context of real-world tasks, and, the influence that factors such as task-load have on task-prioritization performance.

A survey was conducted by Chou, Madhavan et al (1996) on 470 ASRS\(^1\) Incident reports; The survey identified human-errors in 231 reports; According to the surveyors, out of the 231 incident reports, 35 % were associated to errors in task-prioritization, 42% were associated to task-initiation, and 23% to task-termination; They state that high mental-workload was associated to these errors in task-prioritization (Chou, Madhavan, & Funk, 1996). Task-prioritization, task-initiation and task-termination comprise task-management. The term, cockpit task-management is used within the domain of aviation. Cockpit task-management includes task-prioritization, task initiation, monitoring, execution, and termination of multiple tasks in the cockpit.

Colvin, Funk and Braune (2005) conducted two studies to investigate the factors that underlie task-prioritization in CTM. The first study was an initiative to list all possible factors that govern task-prioritization. They recruited 8 airline pilots and had them fly two flight scenarios on the NASA Stone-Stroup Simulator. Both scenarios consisted of 6 events, each of which required making a task-prioritization decision. The experimenters used two methods to interview the pilots on their task-prioritization decisions – one was an intrusive cognitive interview, and another a retrospective interview that accompanied a review of the scenario’s videotaping. Responses generated by the interviews resulted in the identification of 12 factors. They were (1) procedure, (2) status, (3) rate-of-change, (4) needed information, (5) urgency, (6) importance, (7) verifying information, (8) time/effort, (9) salience, (10) consequences, (11) resist forgetting and (12) expectancy. The authors narrowed these down to a subset of six factors: status, procedure, importance, time/effort, urgency, and salience. The second study was conducted to verify if this

\(^1\) Aviation Safety Reporting System
subset of factors relate to task-performance. The second study involved different flight scenarios from the first one. They were low-altitude approach phase of flight that made task-prioritization a critical requirement. Data gathered from the second experiment suggested that procedure was the greatest factor influencing task-prioritization. The experiment which also involved non-intrusive interviews suggested that prioritization is proportional to its, importance, salience, time/effort required to perform it, urgency and inversely proportional to its status. It should be noted that pilot individual-differences may exist in factors that affect their task-prioritization.

Freed (2000) discusses Reactive Prioritization, an optimal prioritization scheme that addresses how tasks can be prioritized in an uncertain environment based on readily available information. The approach, which is intended to be implemented in a computational-agent, tries to minimize the cost of missing deadlines over a given period of time. It takes into consideration four factors for determining if a switch should be made. They are urgency, importance, duration, and interruption-cost. This model, however, is yet to be validated on how people actually make switching decisions, and it does not consider or model the effect of mental-workload on prioritization strategy (Freed, 2000).

Bishara and Funk (2002) studied the effect of training pilots to prioritize their tasks. Three groups of Instrument Flight Rule rated pilots participated in an experiment conducted on a flight-simulator. Each pilot went through a pre-training and post-training session. The first (control) group received no training between the two sessions. The second group received training that included an introduction to Cockpit Task Management (CTM), previous accident reports related to ineffective prioritization, and a summary of factors that lead to CTM errors. The third group received the same as the second, in addition to a prescribed heuristic, aviate-prioritize-execute. Task-prioritization improved in the two training groups, while no such improvement was observed in the control group. However, the improvement may have also been
a result of a learning-effect rather than task-prioritization training (Bishara & Funk, 2002; Parasuraman, 2005).

Raby and Wickens (1994) conducted a study on pilots in a simulated piloting task to explore the strategies used by pilots to adapt to higher workload levels. According to Raby and Wickens (1994), models of task-scheduling some of which specify optimal-performance had been proposed earlier however, very little empirical data existed on how individuals actually scheduled their tasks. Among the previous experimental-studies concerning task-management in pilots, a few indicated that task-prioritization becomes non-optimal at times of high mental-workload. Raby and Wickens (1994) conducted an experimental study with 30 student pilots who performed a simulated landing task. Workload was varied at three levels. The task consisted of various discrete tasks that were assigned one of three priority levels (must, should, can) in addition to the primary continuous task of flight-control. Pilots were observed on how they attended to discrete tasks of varying priority during approach. According to the researchers, the effect of mental-workload was interesting regarding when tasks were initiated. Tasks were not attended to earlier or later during high mental-workload irrespective of priority level, and neither was there an increased optimality in scheduling high-priority tasks. Participants did not show evidence of applying the as-if heuristic on when or how tasks were performed. Under higher workload, they simply increased the shedding of lower priority tasks, consistent with earlier studies. They state that under high workload there are no resources available to make prioritization decisions (i.e., optimality in scheduling was not affected by workload). The time spent on each type of discrete task itself was not affected by mental-workload. Better performing pilots in the experiment more frequently switched between tasks, which is an evidence of switching-flexibility, and they appeared to schedule high-priority tasks earlier. It was also observed that under conditions of high workload, there were more flight path deviations observed within the primary ‘aviate’ continuous task. This deviation exceeded safety limits for poorly performing pilots. Switching over to high-
priority discrete tasks from the high-priority continuous flight-control task may have led to the flight path deviation (Raby & Wickens, 1994).

In an experimental study conducted by Morris and Leung (2006), participants were instructed to perform a manual-tracking task which was the primary task in the experiment. Fifty-two participants were divided into low, medium and high workload groups. Researchers measured flying performance, prioritization errors, and performance within the secondary task(s). Two thirds of the participants (medium and high workload conditions) received a system-gauge monitoring task, that was assigned second priority. One-third of the participants (high workload) received an additional arithmetic task that was assigned third priority. All groups were given a radio-communication listening task that was assigned the last priority for each group. Radio-communication messages were varied in length as a within subjects factor. Message length varied between one and seven chunks. Results of the experiment showed that tracking-error (flying-performance) was significantly different across the (workload) groups. Tracking errors increased with the number of secondary tasks, although the manual-tracking task was instructed to be given first priority. Tracking error was found to be significantly different across message-length, with tracking error increasing as the number of chunks increased. Prioritization errors were also significantly different across the workload groups, as well as across message-length. Generally, prioritization errors were highest on the high-workload condition and lowest on the low-workload condition. The study indicates that individuals can face difficulty in maintaining a prescribed prioritization strategy under conditions of high workload to the point that even the primary task (highest priority) suffers. This is especially important given that the source of increased workload (whether between or within subjects) is associated to a secondary task (Morris & Leung, 2006). The study however did not look at increased workload within the primary task.

Cognitive tunneling is the effect where an operator’s attention is selectively focused on a particular channel of information for a prolonged period of time with other information being left
unattended leading to negative consequences (Hudlicka & McNeese, 2002; Wickens &
Alexander, 2009). Cognitive tunneling can also be seen as the inability to switch onto a secondary
task from a primary task. It can cause operators to miss important cues leading to poor situation
awareness and incorrect decisions or even fatal accidents in the context of car-driving where
information about nearby traffic may be left unattended. This is also closely related to problem-
detection and Inattentional Blindness (Klein et al. 2003).

Iani and Wickens (2007) emphasized the importance of investigating factors that
influence the “compellingness” or “engagement” of displays that support an ongoing-task (OT).
Previous studies suggested that synthetic vision systems (SVSs) or tunnel-displays may tunnel the
operator’s attention, thereby causing operators to miss external alerts or updates. In their study,
the authors explored factors that influence the switching of attention from a primary task to a
secondary task, where the primary task was either supported by a tunnel display or baseline
display. The alert indicating the secondary task, which was a path selection task, was presented
through either the visual or auditory (the more salient one) modality. Additionally, the importance
of the secondary task was also manipulated. The experiment recruited 40 instrument-certified
pilots that flew three simulated approaches. One approach involved a weather update interruption.
Results from the experiment revealed that the frequency of the switching from primary to
secondary tasks was influenced by the modality of the cue as well secondary task importance.
However, increased cue salience did not influence primary task-performance. Interestingly, the
study also revealed that participants in the tunnel display more frequently switched to the
secondary task than in the baseline display. Increased cue salience and the introduction of tunnel
display, showed, improvements in secondary task performance (Iani & Wickens, 2007).

Wickens, Heather and Merlot (1999) conducted a review of 21 experimental studies that
show how display formats, or specific information on displays, can reduce biases that can be
classified as Type-A biases, Type-T biases or a conjunction of the two.
• Type-T calibration bias is an over-reliance on the source of information.

• Type-A calibration bias is a failure to allocate attention or resources to sources of information on the basis of their importance.

Although a number of research studies appear to indicate that operators follow prescribed strategies for attention allocation, research has shown systematic departures from prescribed or optimal scheduling behavior. An underlying factor could be information-overload or high mental-workload. As following an optimal strategy may require making cognitive interpretations of the value of information sources, it places a higher demand on resources of limited capacity, especially when the information-value or importance of cues is not obvious. Type-A biases may be driven by salient cues with lower priority information that may inappropriately indicate a higher level of importance. Sometimes biases can be manifested as Type-A and Type-T conjunctively. They involve an inappropriate allocation of attention (Type-A) made to an information-channel of lower reliability (Type-T). The authors discuss possible factors underlying this type of bias. One factor is a primacy effect where older information may be given more weight than sources of information that may provide recent updates. Other factors are associated to the display. Implicit information on the display such as cueing, 3D information, size, centrality or intensity would make such information perceived to be of higher importance. Explicit information supplementing alerts would provide an estimate of the importance or reliability of alerts may also lead to Type-A and Type-T biases.

Most of the studies reviewed by Wickens, Heather et al. (1999) involved tasks that required some information-integration for good task-performance. The authors reviewed 21 studies involving: calibration, attention-guidance, and display induced calibration. The first category examines the extent to which attention to alarms, information or other sources of information were calibrated with the actual reliability or importance of the information being displayed. The second category examined the effect of attention-guidance, or cueing, on
attention-allocation. The third category concerns provision of displays that incorporate explicit-information indicating the reliability of information and its effect on attention-allocation. The authors of the review provide interpretations and conclusions within each of these categories in the appendix (Wickens, Heather, & Merlot, 1999). Important conclusions from the review are noted below that concern information-integration, divided-attention and the effects of attention-guidance.

Display features that affect information-integration performance:

- Poor reliability has negative effects on information-integration performance.
- Providing explicit information on uncertainty is beneficial to integration performance.
- Implicit reliability indicators are beneficial to information-integration, although explicit indicators could be better.

Display features that affect divided-attention:

- Poor reliability has negative effects on divided-attention.
- Calibration is better when multiple sources of information have equivalent reliability, as opposed to varying reliability (Kerstholt, Passenier, Houttuin, & Schuffel, 1996; Montgomery, 1996).
- Explicit reliability indicators are beneficial to divided-attention.
- Information-sources that aid in prediction may be beneficial to divided-attention (Sorkin, Kantowitz, & Kantowitz, 1988) and indicating their uncertainty can be beneficial to it as well.
Effects of attention-guidance:

- Attention guidance effects on Type-A biases: Research on the role of Helmet-Mounted Displays (HMD) and Hand-Held Devices (HHD) in a Cave Automatic Virtual Environment (CAVE) revealed that the HMD directed attention to simulated targets (tanks, mines) however, this was done at the expense of neglecting a higher priority task involving a nuclear device (Yeh, Wickens, & Seagull, 1998). This effect was mediated when a HHD was used (Yeh, Wickens, & Seagull, 1999). Another study showed similar cognitive-tunneling effects resulting with the use of an immersive display (Thomas & Wickens, 1999).

- Cueing items on the display are beneficial when the cueing is valid or reliable.

- Individuals may over-trust cues with high validity or information-value at the expense of uncued targets (Wickens, Conejo, & Gempler, 1999).

- Cueing items at four levels is better than at two levels.

- Performance slows down as the number of items highlighted increase.

Yeh, Wickens et al. (1999) conducted an experiment in the CAVE virtual reality environment where participants were provided with a Helmet-Mounted Display. Information was provided on far-domain and the near-domain. Participants were supposed to detect, identify and report information about targets in the far-domain while performing a monitoring task in the near-domain. Participants were also supposed to detect a high priority nuclear device that would be presented concurrently with other targets. The nuclear device was not a cued target and was thereby unexpected to the participants. This is supposed to take priority over target detection. The study showed that subjects would report the nuclear device under circumstances it was expected. When expectancies were provided for other targets, target detection improved, however, only at the cost of missing unexpected targets in the environment –**the authors suggest that this was a**
cognitive tunneling effect resulting from the expectancy. The authors state that cueing benefited divided attention, between the near and far domain, however, at the cost of disrupting focused attention in the far domain.

Table 2-1. Summary of literature on task-management

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Major Implications From the Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raby and Wickens</td>
<td>1994</td>
<td>The study showed that pilots followed the prescribed prioritization strategy even under increased levels of workload. Their task-management strategy involved the shedding of lower priority tasks. Pilots did not follow the ‘as-if’ heuristic, and as workload increased, time spent on higher priority tasks increased, and time spent on lower priority tasks reduced. Workload, however, did not have an effect on when sub-tasks were performed – and the optimality on when tasks were performed was not modulated by workload even for the higher priority tasks. Scheduling optimality did not change at increased workload levels. Although the ‘as-if’ heuristic was not followed, under higher workload, for some pilots, the primary task’s flight-path deviations increased due to time spent on some high priority discrete tasks.</td>
</tr>
<tr>
<td>Morris and Leung</td>
<td>2006</td>
<td>The study showed that tracking-error (flying-performance) was significantly different across the (workload) groups, where tracking errors increased with the number of secondary tasks, although the manual-tracking task was instructed to be given first priority. The study indicates that individuals can face difficulty in maintaining a prescribed prioritization strategy under conditions of high workload to the point that even the primary task (highest priority) suffers. The source of increased workload, however, was the secondary task.</td>
</tr>
<tr>
<td>Iani and Wickens</td>
<td>2007</td>
<td>The study explores factors other than workload that may influence the probability that the operator’s attention can get tunnelled. They are the importance of the secondary task, the type of display, and salience of the secondary task cue. The importance of the secondary task can have an effect on the frequency of switching to the secondary task.</td>
</tr>
</tbody>
</table>

Workload Effects on Task-Management

Workload can have an interesting influence on primary task performance. This depends on the importance of the secondary task, the salience of relevant cues within the secondary task,
and the source of the workload. Operators are required to follow a prescribed strategy as seen in the piloting study by Raby and Wickens (1994). Under high-workload, they would shed tasks of lower priority or spend less time on secondary tasks. This is similar to attention-tunneling, however, in task-shedding a critical event or cue from the secondary task is not necessarily missed. Also, task-shedding behavior is a result of awareness about the status of different tasks – tasks of lower priority at the given moment are dropped, within task-shedding. Attention-tunneling, on the other hand, is an effect of mental-workload on task fixation on any given task.

Nevertheless, opportunities to plan ahead can allow operators to allocate attention more optimally. Planning ahead would require operators to develop expectancies. Operator expectancy can be improved by the provision of previews that may aid them to look ahead into the future and accordingly manage high-workload issues more optimally. Previews that provide information about the nature of workload-demand and upcoming threat may alleviate the burden of having to split attention between primary and secondary tasks thereby supporting the prescribed attention-allocation strategy.

The Effect of Previews on Task-Performance

Demands and severity associated to various information channels may vary over time, and although one or more of these channels may be considered primary, high-workload compromises the ability to maintain a prescribed attention-allocation strategy. The previous section discussed break-downs in task-prioritization strategy resulting in attention-allocation behavior characterized by the ‘as-if’ heuristic. The ‘as-if’ heuristic is a result of high-workload where individuals are unable to expect or notice changes in severity within channels not being monitored frequently. Previews or information about events in the near future can aid in the
effective allocation of attention and planning for potential changes in severity. This sub-section
covers Human Factors research that has examined the use of Previews in various real-world tasks.

Segal and Wickens (1990) conducted a study to observe the effects of providing
intelligence or a preview and scheduling-control on a secondary task, on task and mission
performance. The authors particularly wanted to observe operator strategy on whether they could
be categorized as either ‘ants’ or ‘crickets’ (procrastinators). According to Segal and Wickens,
strategy refers to the task-management policy that plays a role in the selection, prioritization and
scheduling of sub-tasks which is an important issue in tasks where conditions change
dynamically. The authors made the following hypotheses – (a) Pilots may display overconfidence
and confirmation bias in selecting strategies. (b) The provision of flight preview information
would lead to more optimal performance. The preview was not displayed but was merely
provided before each flight. (c) When given the flexibility to schedule flight sub-tasks, pilots
would schedule tasks optimally. (d) Stressful conditions would lead to sub-optimal decision-
making behavior.

Segal and Wickens conducted an experimental study with three factors – the availability
of intelligence (or preview), task-difficulty stage in flight (early, late and none), and the freedom
to schedule secondary tasks. Results show that the group with the preview assignment performed
significantly better on all measures of performance. The group given the freedom to temporally
control the scheduling of secondary tasks did not perform significantly better on any of the
performance measures. Preview or Intelligence facilitated task-performance under all conditions
including the no-difficulty condition. However, the provision of the preview did not lead pilots
(within the group that had scheduling flexibility) to scheduling the spatial-side task (the primary
side task) under phases of flight with lower task-difficulty. However, it facilitated the scheduling
of the checklist task (secondary side task) at more convenient times and it had an indirect benefit
to fixed-time tasks as it eliminated the need to monitor other tasks when it was known that a
difficulty was not coming up.

Segal and Wickens conclude that without intelligence operator behavior was closer to
that of ‘crickets’ rather than ‘ants’ where deprived of information about future events, operators
did not invest resources to prepare for ‘lean times’ and were overconfident about the assumption
that the future would have ample time for task-performance. In other words, operators, without
the provision of intelligence were not prepared for the ‘worst-case scenario’. Intelligence about
difficulties in flight supported performance at both the sub-task and mission level (Segal &
Wickens, 1990).

Andre, Heers and Cashion (1995) studied the effect of declarative and procedural
previews in piloting. Research within a variety of domains has explored the role of human
strategic behavior (Adams, Tenney, & Pew, 1998). Workload management or the scheduling of
various sub-tasks within a larger mission is an important aspect of strategic behavior, and
according to the authors, little is known about how specific factors influence scheduling behavior
or task-prioritization. Their experimental study explored the effect of providing a workload-
preview or heads-up, on task-prioritization.

There were two phases of aircraft-piloting and three secondary tasks. The first phase was
the low-workload condition, and the second phase was the high-workload condition generated by
a disorganized instrument-layout and increased turbulence. Before the two phases, pilots went
through a practice session. There were three workload preview groups (no preview, declarative
preview, procedural preview), and six pilots were assigned to each of them on a random basis. In
the declarative and procedural preview groups, pilots were warned about nature of the workload
condition beforehand. Only the pilots in the procedural group were given practice in the high-
workload condition. Although, not shown to be statistically significant (F(2,30) = 1.62), pilots in
the two preview groups scheduled more secondary tasks during low-workload than during high-
workload. In the no-preview condition, pilots scheduled secondary tasks equivalently across the two conditions of flight. An analysis of a particularly difficult secondary task – target-acquisition, revealed a significant interaction effect (F(2,30) = 4.21, p < .05) between flight workload and workload-preview conditions, where, the no-preview group scheduled the secondary task twice as many times during the high-workload condition than in the low-workload condition. The preview groups, however, scheduled fewer target-acquisitions during high-workload compared to low-workload. Andre, Heers and Cashion (1995) state that pilots did not dynamically schedule tasks and that this was done ‘in streaks within each flight phase’. In terms of performance, the procedural preview group was superior on overall flight performance followed by the declarative preview group, followed by the no-preview group. The procedural preview group was superior to the other two groups in flight-performance and scheduling strategies. Overall, the study demonstrates the benefit of providing a workload-preview in scheduling strategy (Andre, et al., 1995).

Wickens, Pizzaro and Bell (1991) investigated the effect of a preview in a decision-making task. Although the task involved manual-control in piloting an aircraft, the preview information was intended to optimize the decision-making component of the task concerning navigation. The preview here was intended to aid the decision-making process by aiding participants to ‘look into the future’ or globally optimize their decisions. According to the authors, individuals do not take into consideration future consequences in decision-making, and they appear to be more reactive rather than proactive in their responses. There is little agreement on the usefulness of previews beyond the manual control domain (Wickens, et al., 1991). An experimental-study reported by Tulga and Sheridan (1980) involving a resource-allocation task found that even the provision of a preview led to participants allocating their attention based on immediately foreseeable events, and their ability to plan into the future declined as the number of attention-channels increased (Tulga & Sheridan, 1980).
The task in the experiment was a flight-control and decision-making task. Locations within the simulated terrain varied in the number of casualties present and the score depended on the number that could be rescued. Participants would fly a helicopter from node to node, and at each node, they were required to decide which node to fly to next (out of two options). The score would increase with the number of casualties rescued and would decrease with the distance traveled. Turbulence varied at four levels at each node-to-node path, and higher turbulence would correlate to a smaller probability of traversing between those nodes.

The main within-subjects factor was the preview. The preview condition provided the entire graph-map including turbulence-levels, casualties and distances between all nodes. This preview was intended to aid the decision at each node and make participants consider the global decision space. The non-preview condition provided the variables that pertained to traversing to the very next nodes which restricted their decision making to the local space. The results of the experiment showed that decision-making performance was significantly worse in the preview condition. The authors conducted a strategy-analysis to analyze which variable was optimized the most by the participants among casualties, turbulence, value, score and distance. The analysis showed that in the non-preview condition - participants tried to locally maximize score as often as maximizing casualties; however, in the preview condition – participants attempted to maximize casualties and did not appear to maximize score that required the consideration of probabilistic information along with other variables. The authors also indicate that decisions involving long-term information considered in the preview-group were not significantly different from short-term strategies in influencing performance. They conclude that the preview overloaded the participants to the extent that they could not integrate the probabilistic information required to successfully optimize their decision making at the global-level. This failure to integrate information correctly led to a decision-making strategy poorer than the one used by merely considering local variables in the non-preview condition (Wickens, et al., 1991).
Cummings and Mitchell (2005) investigated the effect of Level-of-Automation (LOA) on multiple Unmanned Aerial Vehicle (UAV) task-performance and supervisor situation awareness. The LOA was varied in the mission and payload management aspect of the task that required the scheduling of targets. Navigation and UAV control was fully automated, and the LOA was not varied within these aspects. The third LOA (active automation) in this task consisted of a workload-preview that would specifically indicate upcoming high-workload periods resulting from simultaneous and conflicting tasks within mission. There were significant differences in performance levels only in the high re-planning level. Within high re-planning, the level-3 automation condition consisting of the workload-preview resulted in the lowest task-performance among the four LOA conditions. The authors expected that the workload-preview would aid participants to plan ahead of time. The result of the workload-preview, however, was that participants re-planned their targets (using a request time-on-target or TOT change) more often than they should have, despite that this flexibility (in the high re-planning condition) was instructed to be used sparingly. Operators attempted to globally optimize their schedule through re-planning actions but could not do so due to high mental-workload. The difficulty in integrating probabilistic information into scheduling made the use of workload-preview even harder (Cummings & Mitchell, 2005).

Metzger and Parasuraman (2005) examined the effect of an air-traffic conflict-detection alerting system (a predictive aid) on task-performance and mental-workload in an air traffic control task. The authors conducted two separate experiments with perfect and imperfect conflict alerting systems. In the first experiment with perfect prediction, the primary task was the detection of potential conflicts. The secondary task constituted communicating with pilots to hand them off to sectors. The experiment was 2X2 within subjects design involving conditions - with predictive automation and without predictive automation, moderate task-load and high task-load. The experiment measured task-performance in terms of the percentage of conflict detections, self-
separation detections, and the response times of conflict-detection and self-separation notifications; and mental-workload measured by NASA-TLX. The alerting system under perfect reliability was beneficial to performance as detected self-separations were higher in percentage when the aid was present. The presence of the aid was also associated with a higher percentage of hand-offs. For aircraft handoffs specifically, an interaction effect between task-load and the presence of the aid was observed – the effect of the aid was more pronounced under high task-load. There was a greater tendency to offer premature hand offs without the predictive aid especially in the high task-load condition. The authors report that mental-workload was not reduced with the predictive aid, as freed mental resources in the presence of the aid were used towards the secondary task. In the second experiment, the presence of the conflict-alerting aid with imperfect reliability degraded task-performance (Metzger & Parasuraman, 2005).

Table 2-2. Summary of literature on workload-previews

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Implications of Incorporating Previews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings and Mitchell</td>
<td>2005</td>
<td>The preview was detrimental to performance as operators attempted to globally optimize their schedule each time a ‘bottleneck’ was predicted by the automation. This attempt to globally optimize the schedule with a re-planning operator (an affordance) resulted in poorer performance in comparison to other levels of automation without the preview, or in the same level of automation, in another condition with significantly lesser re-planning flexibility.</td>
</tr>
<tr>
<td>Wickens, Pizzaro and Bell</td>
<td>1991</td>
<td>The preview was detrimental to performance. The preview that consisted of information to facilitate long-term planning overloaded participants to the extent that they could not integrate the probabilistic information required to successfully optimize their decision making at the global-level. The non-preview condition that was not provided with global information performed better through the application of a local optimization strategy.</td>
</tr>
<tr>
<td>Segal and Wickens</td>
<td>1990</td>
<td>The preview was successful in facilitating performance. Without the preview, operator behavior was closer to that of ‘crickets’ rather than ‘ants’ where deprived of information about future events, they did not invest resources to prepare for ‘lean times’ and were overconfident in their ability to address tasks in the future. In other words, operators, without the provision of the preview were not prepared for the ‘worst-case scenario’. Preview in the form of forecast information about future difficulties in flight supported performance at both the sub-task and mission level. The preview, however, was static and was provided prior to flight.</td>
</tr>
<tr>
<td>Andre, Heers and Cashion</td>
<td>1995</td>
<td>The preview facilitated performance. Pilots in the preview groups scheduled more secondary tasks during low-workload than during high-workload. In the non-preview condition, pilots scheduled secondary tasks equivalently across the two conditions of flight. The analysis of the scheduling of a difficult secondary task showed that the no-preview group scheduled the secondary task twice as many times during the high-workload condition than in the low-workload condition. The preview groups, however, scheduled fewer of those tasks during high-workload than during low-workload. The preview was provided once before simulated flight, and it facilitated the scheduling of secondary tasks in preparation of a high workload phase during flight.</td>
</tr>
<tr>
<td>Metzger and Parasuraman</td>
<td>2005</td>
<td>The preview facilitated performance. The study objective was to investigate if automation reduces mental-workload in a Free-Flight Air-Traffic Control task and improves performance. Additionally, they also looked at the effect of automation imperfection on measures of mental-workload and task-performance. The preview/automation provided a 6 min look-ahead of a potential air traffic conflict. The conflict-detection automation with perfect reliability facilitated performance, however, imperfect conflict-detection did not facilitate performance.</td>
</tr>
</tbody>
</table>

**Discussion of Previous Literature on Previews**

Previous studies on previews incorporated in various tasks have shown mixed findings on its usefulness. The literature has shown that the provision of static preview information prior to simulated flight is beneficial to pilots in terms of their ability to schedule tasks effectively to have sufficient resources to maintain focused attention during the high-workload phase of flight (Andre, et al., 1995; Segal & Wickens, 1990). A study that examined the effect of conflict-detection automation (a type of dynamic preview) revealed that the preview facilitated performance under perfect reliability; however, the preview with imperfect reliability did not improve performance. Studies by Wickens, Pizzaro and Bell (1991) and Cummings and Mitchell (2005) have shown that previews with probabilistic information may not be useful in planning ahead or in making decisions to optimize a global variable, as the global decision-space of possible actions becomes too complex to mentally process under high task-load. This relates to the notion that individuals make decisions in accordance to a satisficing criterion rather than optimizing a variable (Simon, 1959).
To generalize across different cases, it appears that preview information that contributed to cognitive-load was detrimental to performance. Increased cognitive-load resulting from preview information may have been associated to imperfect preview reliability, probabilistic information or complex display formats. Parasuraman (2005) provides a review of issues pertaining to task-prioritization. Under high mental-workload, operators may switch into a more sequential than concurrent strategy and task-prioritization suffers. Although tasks of higher-priority should be given attention over low-priority ones, some low-priority tasks can be attended to and completed if they take significantly less time, however, individuals may tend to shed the low priority tasks altogether. Parasuraman (2005) points that more research is required to be done to study the operators ability to dynamically schedule tasks on the fly which is pertinent in environments where workload-demands may vary or where surges in workload may appear unpredictably. This dissertation focuses on the effect of providing a real-time comparison of severities associated to different sub-tasks in cyber-security event monitoring: It is hypothesized that predictive information about upcoming events and their corresponding severity levels may facilitate task-prioritization and timely task-switching upon unexpected changes in sub-task priorities. If so, this would bring insight in the ability to dynamically schedule tasks and how this ability is affected by a workload-preview under different conditions of task-load.

All previous studies involving previews have considered the aviation or command-and-control domain. The cyber-security event-monitoring task considered here has different cognitive demand characteristics. The workload-preview considered in this study differs from the static forecast preview information examined in studies reviewed here (Segal and Wickens, 1990, Andre, et al. 1995). Although previews with dynamic predictive information are reported, the dynamic workload-preview in the multiple-UAV management task resulted in poorer task-performance and scheduling performance (Cummings and Mitchell, 2005); In the other study concerning the effect of collision detection, the collision-detection predictor improved primary
task performance, but the study did not address task-prioritization or scheduling (Metzger and Parasuraman, 2005).

**Integration and Hypotheses**

The literature review covered various theoretical frameworks and experimental-studies on attention, situation awareness, mental-workload, interruptions, task-prioritization, the effects of workload on task-prioritization and task-performance, and the effects of previews in various real-world tasks. Research on previews has mainly considered the aviation and command-and-control domain, and, experimental research on its effect on task-prioritization has mainly focused on tasks in which sub-task priorities remain constant. Most research conducted on interruptions, specifically on the effect of warnings and interruption-lag, have focused on the dependent-variable of resumption-lag (i.e., the time taken to resume a primary task). Moreover, unlike previous research on interruptions relevant to this dissertation, the task being considered in this dissertation does not distinguish sub-tasks as interrupting versus ongoing tasks. The sub-tasks considered within the scaled-world task incorporated in this study mainly differ in default priority associated to them. The current study primarily investigates the effect of workload-previews on task-prioritization and task-performance. Specifically, the workload-preview displays information related to sub-task severities and event-volume. Together, these components of the workload-preview provide an estimate of time-pressure associated to individual sub-tasks. The workload-preview is embedded in a dual-task event-monitoring context. In this scaled-world cyber-security task, changes in priority associated to various sub-tasks are unexpected. These changes in priority will be incorporated in the form of surge-events within the secondary task. Surge-events in the secondary task will be assigned priority over concurrent primary task events.
**Research-Question** – In dual-task event monitoring for a cyber security application, does the provision of a Workload-Preview improve Task-Performance and Task-Prioritization and to what extent does improvement differ between Moderate and High Task-Load levels?

High Task-Performance requires attending to event descriptions as soon as they arrive to allow enough time to comprehend event descriptions. Task-Performance can be measured by response-time and accuracy. Successful task-prioritization is dependent on the awareness of relative sub-task priorities. A response to an event within a sub-task of lower priority, while concurrent events of higher priority are active, is a task-prioritization error. A task-prioritization error can be a result of not detecting changes in overall severities associated to individual sub-tasks displayed. The simulated task consists of surge-events and non-surge-events.

In the case of surge-events, the Workload-Preview provides a larger window of opportunity (Miller, 2002) to notice unexpected changes in severity-level. In the default case, the primary task will have severity greater than or equal to that of the secondary task. In the non-default case or during the presence of surge-events in the secondary task, the secondary task’s severity will exceed that of the primary task. The Workload-Preview provides a look-ahead on changes in severity-level in the secondary task and thereby provides a larger window of opportunity to develop awareness about changes in sub-task priorities. The larger window of opportunity will not be effective if participants in either condition (i.e., with and without Workload-Preview) do not scan the display frequently enough to notice changes in severity. According to literature reviewed, Workload-Previews have been beneficial to task-prioritization in a variety of tasks especially when the Workload-Preview is purely reliable (Andre, et al., 1995; Segal & Wickens, 1990). It is anticipated that the ability to notice surge-events in a timely manner will benefit Task-Prioritization as the prioritization of surge-events over non-surge-events should reflect on the measure of Task-Prioritization (H3).
From a perspective of ecological theory of social perception, the preview would be associated to more dynamic events (i.e. changes in Overall-Severity as they appear in additional points in time) that would supplement information from an otherwise more static display, that would facilitate perceptual noticing (McArthur & Baron, 1983). The larger window of opportunity for detecting changes in severity levels would potentially reduce mental-workload. In the non Workload-Preview condition, the requirement to frequently scan and process severity information on the display to check for relative changes in priority will contribute to mental-workload (Mark, et al., 2008). This form of interference or interruption would result in delayed response times on events and thereby poorer Task-Performance in the Non Workload-Preview condition (H1); furthermore, as the simulated task would involve attending to several batches of events; responses to most secondary task events would be temporally placed between responses to primary task events; switch times on switches from secondary task events to primary task events when no surge-events are present will be expedited with the provision of the Workload-Preview. In this case, primary task events will be prioritized over secondary task events in a timely manner, thereby reflecting both Task-Prioritization (H3) and Task-Performance (H1).

Literature has shown that operator attention-allocation follows the as-if heuristic under high Task-Load (Morris & Leung, 2006; Raby & Wickens, 1994). The above discusses the benefit that the Workload-Preview would bring under Moderate Task-Load. In the High Task-Load condition the availability of a larger window of opportunity to take notice of changes in priority would be more beneficial to prioritizing surge-events over non-surge events which would reflect on overall Task-Prioritization (H4). As the simulated task presents events in batches, attending to secondary task events would be temporally placed between attending to primary task events with the exception of cases where surge-events are present. Switching from the secondary task to primary task events in a timely manner would positively reflect on Task-Prioritization. However, there would be a higher probability of failing to do so under high time-pressure in
comparison to that of the Moderate Task-Load condition. Therefore, the benefit of the Workload-Preview on Task-Prioritization would be more pronounced under High Task-Load (H4).

In the non Workload-Preview condition, as discussed above, the requirement to frequently scan for changes in task priorities would contribute to higher mental-workload resulting in poorer Task-Performance in terms of response-times and successfully answering events (Mark, et al., 2008). The presence of a Workload-Preview would mitigate disruptive effects of frequently scanning for severity-information. As interruptions can be more disruptive under high Task-Load (Iqbal & Bailey, 2005), this mitigating effect of the Workload-Preview would be more pronounced under High Task-Load. In other words, under high time-pressure the requirement to frequently scan for severity information on the display would be costlier in comparison to that in the moderate Task-Load condition. Therefore, the Workload-Preview will be more beneficial to Task-Performance in the High Task-Load condition (H2). To summarize, the hypotheses are stated as follows:

**Hypotheses**

H1 - The Workload-Preview will improve Task-Performance.

H2 - Task-Performance improvement resulting from the Workload-Preview will be more pronounced in the High-Workload condition.

H3 - The Workload-Preview will improve Task-Prioritization.

H4 - Task-Prioritization improvement resulting from the Workload-Preview will be more pronounced in the High-Workload condition.
Chapter 3

Materials and Methods

The effect of a technological aid on human Task-Performance varies with the task and context within which it is being incorporated. This study focuses on a decision-making task similar to that of a cyber-security analyst monitoring an organizational network. It specifically focuses on part of the task that involves monitoring events within the organizational network and assigning them a category and priority level. The scaled-world task that emulates organizational cyber-security monitoring used in this study is called NETS-DART. This chapter provides a description of NETS-DART, the design of the Workload-Preview, the experimental design, scenario design, event design and experimental protocol.

The Cyber-Security Dual-Task Context

The scaled-world task used in this study is named NETS-DART. NETS-DART is an offshoot of idsNETS (Mancuso, Minotra, Giacobe, McNeese, & Tyworth, 2012), which itself is a modification of the NeoCities 3.1 simulation (Hamilton, et al., 2010). DART is the short form for Dual-Task Attention Research Testbed. The task consists of an organizational setting consisting of several departments. Each department is categorized as a core department or an external department. The participant is required to monitor events in a hypothetical organizational network with events embedded within various departments. Several events can appear concurrently within different departments. The task requires the participant to use an appropriate task-prioritization strategy to respond to these events. In this task, the core departments (e.g., Accounting, Engineering) are considered important to organizational functioning and the external departments
(e.g., Recreation Center, Warehouse) are considered less critical. Monitoring and responding to events in the core and external departments are the primary and secondary tasks, respectively, in the scaled-world task. These will be referred to as primary and secondary tasks here. The rule knowledge required for responding to events in primary and secondary tasks are identical, however, events within these sub-tasks may only differ in priority. The rule knowledge trained to participants pertains to responding with, one of four category-assignments (listed in next paragraph) and one of three priority-level assignments (listed in next paragraph). All possible category-assignments and priority-level assignments should be considered as possible options to answering an event irrespective of whether the event appears in the primary task or secondary task.

An event is a potential threat in the organizational network which is presented on the interface in the form of a report description. An event possesses several properties such as a time-limit, priority-level, category and a dispatch-time when it would appear during the length of the scenario. Figure 3-1 shows a sequence of states (or sub-events) that proceed in any given event in NETS-DART. This structure inherits many properties of a typical NeoCities event, however, the NeoCities resource-allocation concept has been replaced by a task function of priority level assignment. Upon reading the event description, a response to the event constitutes category and priority-level assignment. The scaled-world task includes four types of event categorizations – Virus, Worm, Vandalism, and, Espionage. Events of each category would be associated to one of three priority-levels. Priority-levels 1, 2 and 3 represent an increasing order of importance.
Figure 3-1. Sequence of states within an event

An event description is necessary and sufficient to determine category and priority-level. Once these are assigned, a feedback message is provided within a fixed delay interval. The feedback indicates whether the categorization was correct or incorrect, however, it does not provide information on the correctness of assigned priority-level. Immediately after a correct categorization, an event may continue to remain active. The duration of time for which an event remains active is determined by the response-time, assigned priority-level, and, the magnitude of the event at the time of the response. An event does not remain active beyond its time-limit. The mapping incorporated here between event magnitude and priority-level is provided in the next section. The simulation allows for multiple attempts to respond to a given event, however, this would not result in the same score from a potentially earlier response with correct assignment of priority and category. It is important to note that scenarios designed for this experiment, consist of several batches or groups of events that are temporally separated (i.e., all the events in a scenario are not presented all at once in the beginning).

Figure 3-2 provides the screen capture of the NETS-DART interface provided in one of the two experimental conditions. Each panel in the interface is labeled. The Department Monitor in Panel A is used for selecting a specific department. On the top of Panel A, a button is provided that
affords switching between external and internal departments. Departments from only one of the two groups would be visible at any point in time. A department that consists of an active event would be accompanied with a filled green circle on the left. Upon clicking a specific department, Panels B and C are refreshed to provide information specific to the selected department. Panel B is named ‘Report Description’, and it provides the event description if an event is present in the selected department. For scenarios in this experiment, only one event description is present at a given time as multiple events within the same department do not occur concurrently. When an event completes, it is removed from the Report Description panel.

Figure 3-2. Screen capture of the NETS-DART interface

Panel E consists of the features required for responding to an event with category and priority-level selections. Responses cannot be made unless both category and priority-levels are assigned.
Panel C provides feedback about responses made and feedback when an event completes. Panel D is the Task-State Overview, and this consists of the Workload-Preview present in only one of the experimental conditions. The Task-State Overview provides the necessary information about when to switch between primary and secondary tasks. This is described in more detail later in this chapter.

**Task-Prioritization, Event Magnitudes and Event Priority-Levels**

The previous section introduced the NETS-DART interface and the features used for interacting with the task. This section covers task-prioritization in NETS-DART, how it ties to Overall-Severity, event magnitudes and priority-level assignments. NETS-DART provides a simulation of a dual-task environment where analysts may be occasionally presented with unexpected workload-surges within a secondary task. The external and internal departments each (i.e., primary and secondary tasks) have an Overall-Severity associated with them. Participants are required to give higher priority to the sub-task with the higher Overall-Severity. As the Overall-Severity may change unexpectedly in a scenario, the task requires immediate switching to the sub-task with higher Overall-Severity, and responding to events within that sub-task. If the Overall-Severities are equal, the primary task or the core-departments should be given priority over the secondary task. The measure of Task-Prioritization incorporated in the study aims at capturing the extent to which this Task-Prioritization rule is followed.

In the simulation interface, Overall-Severity information is provided in the Task-State Overview discussed in the next section. As mentioned earlier, an active event has a magnitude. Overall-Severity for a given sub-task is the average of magnitudes of active events within it. An event’s magnitude would either remain fixed for its entire length if the event is not addressed (i.e. category and priority), or it would be allowed to grow. Most events in the scenarios designed for
this experiment have a magnitude cap equal to their initial magnitudes (i.e., they would not grow in magnitude), the exception being surge-events present in surge scenarios. The event’s initial magnitude is pre-set in the scenario description and corresponds directly to its priority-level, with the exception to surge-events. Table 3.1 provides the mapping between the priority-level of the event and the event’s magnitude cap.

Table 3.1. Mapping between event priority-levels and event magnitude-caps

<table>
<thead>
<tr>
<th>Event Priority Level</th>
<th>Event Magnitude Cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
</tr>
</tbody>
</table>

For surge-events, magnitude levels grow based on the magnitude growth model inherited from CITIES (A. Wellens & Ergener, 1988). An event with growing magnitude would be associated with increased time-pressure, in comparison to an event with a comparable initial magnitude that remains constant. Figure 3-3 illustrates the characteristic of the CITIES magnitude growth model. The curve representing an event with an initial magnitude of 5.0, reaches its magnitude-cap set at 8.0 after which its magnitude remains constant. The magnitude growth model is also used for non-surge-events although magnitudes do not grow in those cases – they shrink according to the magnitude growth model when correct categorization is applied (not illustrated in Figure 3-3).
A modification to the scoring model used in the NeoCities was made to accommodate the changes in the context and the type of response that is required to be made – category and priority assignment. The original NeoCities involved crisis management and assignment of fire, police, and hazmat resources. An event would require a resource or unit to resolve (e.g. squad car). Moreover, the magnitude of an event in NeoCities would correspond to a minimal number of units required to be assigned within a given amount of time to successfully complete the event. In this simulation, the ‘minimal number of units’ requirement is replaced by task function of priority-level assignment. Assigning a priority-level equal to that of the event-priority would result in a higher score than over-prioritizing or under-prioritizing the event. Table 3-2 provides the translation of the assigned priority-level to the number of ‘resources’ that would be applied in the magnitude-growth model incorporated in this study. In NETS-DART, successful event completion is only conditional upon whether a correct category was applied to the event. Upon correct categorization, under-prioritization or over-prioritization will only result in a lower score – not event failure. In NeoCities, successful event completion would require correct resource
application and ‘timely’ resource application with a sufficient number of units (i.e., magnitude should shrink to zero before event time-limit is reached).

Table 3-2. Translation of assigned priority-level into the number of resources applied

<table>
<thead>
<tr>
<th>Assigned Priority-Level</th>
<th>Resources Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Event-Priority-Level</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

**Design of the Task-State Overview**

The previous section stated that sub-tasks have Overall-Severities used for comparing priorities of sub-tasks. This section describes the Task-State Overview as it applies to the task and the design decisions made for the Task-State Overview. The Task-State Overview consists of a Workload-Overview and a Workload-Preview. Figure 3-4 illustrates the planned design of the Task-State Overview including the Workload-Preview component. The Task-State Overview has an upper section and a lower section. Both provide information about active events within each sub-task.
Figure 3-4. Initial design of the Task-State Overview including the Workload-Preview

The upper section of the Task-State Overview provides a summary of Overall-Severity providing task-priority comparison between primary and secondary tasks. The lower section provides comparison on the number of events (or reports) between primary and secondary tasks.

Severity information and the number of events together provide an estimate of time-pressure and workload associated within each sub-task. For example, when the two sub-tasks have equal Overall-Severity, the sub-task with more events within it would be associated to higher time-pressure. On the other hand, if the two sub-tasks have an equal number of events, and if their current Overall-Severity are comparable, a sub-task with an increasing Overall-Severity
with time would be associated to have higher time-pressure in comparison to a sub-task with constant Overall-Severity.

Accordingly, the predictive information provided is termed the Workload-Preview. The Workload-Overview includes information about past state. Past state information is provided to aid in the perception of state changes. While the comparison of current Overall-Severities is sufficient to trigger a switch from one sub-task to another, the provision of past state information would provide information about extent to which Overall-Severities and event numbers have changed over the last 6 to 12 seconds. This could aid in faster task-switching in cases where a switch is intentionally delayed during an ongoing event response.

Overall-Severity is accompanied with, lowest and highest magnitudes within a given sub-task. These are represented in a candlestick plot (a line and a filled rectangle shown in Figure 3-4) format that is similar to a Box-Plot, although less detailed than a Box-Plot. The Overall-Severity is represented by the filled rectangle. The Task-State Overview does not take into consideration active events in the task that have been answered correctly by the participant as it aids in responding to unattended events. The primary and secondary task information has been provided within the same plot area, as opposed to separate plot areas, as the former design would have a reduced visual-scan cost thereby facilitating comparison. The refresh-rate of the workload preview and overview is set at six seconds (time-step in the simulation has a length of three seconds). The maximum foresight the Workload-Preview provides is twelve seconds – it provides future state six seconds ahead of time, and twelve seconds ahead of time. The preview includes events that are yet to be initiated. The primary task uses colors of saturation higher than that of the secondary task for all variables as higher saturation appears to be more salient, thereby, intuitively suggesting the primary task.
Experimental Design

Table 3-3 summarizes the main independent variables in the study, and Table 3-4 summarizes the main dependent variables in the study. Additional independent and dependent variables are provided in Table 3-5 and Table 3-6. The following sub-sections explain their implementation in more detail.

Table 3-3. Main Independent Variables

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Number of Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task-Load (Within-Subjects)</td>
<td>2 (Moderate vs. High)</td>
<td>The number of events that appear in a given duration of time is increased to generate higher Task-Load</td>
</tr>
<tr>
<td>Preview-Availability (Between-Subjects)</td>
<td>2 (Presence vs. Absence)</td>
<td>Provision of predictive information on Overall-Severity and event numbers.</td>
</tr>
</tbody>
</table>

Table 3-4. Main Dependent Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task-Performance</td>
<td>Normal-Score</td>
<td>It is calculated based on categorization accuracy, prioritization accuracy, and, response time</td>
</tr>
<tr>
<td>Task-Prioritization</td>
<td>Task-Prioritization Error</td>
<td>It is the number of times a deviation was made from a prescribed sequence of event response</td>
</tr>
</tbody>
</table>

Table 3-5. Additional Independent Variable

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Number of Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surge-Presence (Within-Subjects)</td>
<td>2 (Presence vs. Absence)</td>
<td>The presence of secondary task events that grow in severity over time and exceed that of concurrent primary task events</td>
</tr>
</tbody>
</table>

Table 3-6. Additional Dependent Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surge-Prioritization</td>
<td>Surge-Prioritization Error</td>
<td>This is similar to Task-Prioritization Error, but, specific to surge-events.</td>
</tr>
<tr>
<td>Subjective-Workload</td>
<td>TLX-Score</td>
<td>It is a modification to NASA-TLX</td>
</tr>
</tbody>
</table>

Subjective-Workload
Implementation of Independent-Variables

**Task-Load** is manipulated by varying the number of events presented to a participant in a given period of time. Higher Task-Load is associated with more events within a given scenario than Moderate Task-Load within a scenario of same length. Task-Load was implemented by varying time pressure between scenarios. Time-Pressure is a dimension of workload and, Metzger and Parasuraman (2005) in a study involving a conflict detection task varied task-load in 30 minute scenarios by manipulating the number of aircraft within these scenarios. Results from the pilot-test conducted before conducting the main experiment for this dissertation, the time-pressure manipulation was successful in achieving differences in subjective-workload. More details on this manipulation are provided in the scenario design and event design section.

**Surge-Presence** represents the presence or absence of surges in a given scenario. Scenario C consists of two secondary task events that increase in magnitude with time. These are referred to as surge-events. Scenario D consists of 3 surge-events. Scenarios A and B are baseline scenarios that do not consist of any surge events. Results from the pilot-test conducted before conducting the main experiment for this dissertation, indicated that, at both levels of Task-Load scenarios with surge-events were associated with higher Task-Prioritization Error in comparison to scenarios without surge-events although this difference was not statistically significant.

**Preview-Availability** is a between-subjects factor that represents the type of interface that the participant interacts with. Figure 3-5 provides a screenshot of the Task-State Overview component in the client interface. The region surrounded with the dashed green line is the Workload-Peview. Depending on this experimental condition, a participant would either receive the whole Task-State Overview or the portion of the Task-State Overview excluding the
Workload-Preview. The upper section of the display (the rectangles) provides Overall-Severit
information while the lower section of the display (triangles) provides information on event-
volume or the number of reports as indicated to participants. The primary-task or the ‘core-
departments' are indicated by more saturated colors (e.g. bright blue or bright red are indicative
of the primary-task). The upper section displaying Overall-Severity and the lower section
displaying event-volume, both, provide past and current states. The Task-State Overview does not
display history over twelve seconds old. The current state is updated every six seconds and at
every update the current information is passed over to the section of the panel providing past
information – this involves updating the six second old information. Also, at every update the six
second old information is passed over to section of the display providing twelve second old
information.

Figure 3-5. Task-State Overview component of the client interface including Workload-Preview
Implementation of Dependent-Variables

**Normal-Score** is a measure of Task-Performance and its value if based on response-times and accuracy of assigned category and assigned priority-level. The computation of the Normal-Score is based on previous implementations of NeoCities with a modification – the assignment of resource numbers is replaced with a specific value obtained from Table 3-2 provides the mapping between the assigned priority-level and the resources applied in the magnitude- growth equation. Normal-Score for a scenario is the average of Normal-Scores for each event within the scenario – this could be referred to as *Event Normal Score*. For any event in the simulation, the simulation computes the *Response Damage Area* resulting from the response to that event. The calculation of the *Event Normal Score* is determined by the *Response Damage Area* and it is compared with the *Worst Case Damage Area* and *Best Case Damage Area*. The formula for *Event Normal Score* is as follows:

\[
\text{Event Normal Score} = 100 \left( \frac{\text{Worst Case Damage Area} - \text{Response Damage Area}}{\text{Worst Case Damage Area} - \text{Best Case Damage Area} + 1} \right)
\]

The computation of **Task-Prioritization Error** is based on the number of sequencing errors made in a given scenario. Every scenario designed for the experiment has an associated prescribed order in which events should be attended to. The prescribed order is an order of event responses that would be obtained if Overall-Severity information was used to guide the selection and responses to events embedded within primary and secondary sub-tasks. The prescribed order can be expressed in the form of predecessor lists. Each event may or may not have a predecessor list associated to it. As an example, if an event E2 has predecessor list including E1 and E3, the prescribed order to attend to the events would be – E1, E3, E2 or E3, E1, E2. If an event response sequence is in the order E1,E2,E3; this is counted as1 sequencing error, because E2 was attended to before its predecessor E3. If an event response sequence is in the order E2, E1, E3, this is
counted as 2 sequencing errors as both predecessors do not appear before the response E2. The sequence of event responses is obtained from action history information recorded within the output obtained at the end of each scenario performed. This is used to obtain sequencing errors. A script is developed to read the file and generate the number of sequencing errors. It should be noted that, for scenarios developed here, some predecessor lists are conditional upon event response-times, as events intersperse and prescribed priorities for secondary task events can shift after pre-set time deadlines. Each scenario has a maximum number of sequencing errors associated to it. It should be noted that Task-Prioritization Error does not depend on whether events were assigned correct categories or priority-level assignments. It is possible to obtain zero Task-Prioritization Errors (perfect prioritization), and yet obtain a 0 for Normal-Scores associated to each event.

**TLX-Score** is a measure of Subjective-Workload, and it uses the same dimensions as NASA-TLX. However, its implementation deviates from NASA-TLX due to limitations present in the online survey tool incorporated in this study. TLX-Score is obtained by taking the sum of 5 dimensions – mental-demand, physical-demand, temporal-demand, effort, and frustration. Participants provide their subjective estimation of each of these variables on a 1 to 21 scale. The performance dimension was dropped for the calculation of TLX-Score as the end of each scenario provides participants with partial information on how they performed in a performance report update. This could have confounded subjective performance ratings.

**Scenario Design and Event Design**

Scenario design implements the requirements of the experimental design and the term mainly refers to characteristics of scenarios such as duration, number of events, concurrency of events,
and the presence of specific conditions such as surge-events within the scenario. The scenario
design is related to event design. Event design refers to the specific characteristics of events such
as event descriptions, number or type of event categories, priority-levels, and, the underlying cues
placed for indicating event categories and priority-levels. The experiment consists of a training
scenario and four performance scenarios. Table 3-7 lists the characteristics of each scenario. This
table indicates that Scenario’s C and D are designed for increased time-pressure with more events
compressed within nearly the same amount of time as Scenarios A and C. Therefore, scenarios B
and D can be labeled as High Task-Load scenarios. The number of events for Moderate and High
Task-Load were chosen based on previous experiments with NeoCities and this scenario-design
was tested in a pilot-test that showed significant differences in subjective-workload between
Moderate and High levels of Task-Load. Scenarios C and D are scenarios that consist of surge-
events. The objectives of the study entail exploring the human ability to notice and respond to
surge-events within the secondary task. However, scenarios that consist of surge-events are
required to be compared with scenarios without surge-events. Within each scenario, events were
separated in batches, and three events were initiated within the same batch.

Table 3-7. Overview of scenario design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>No. of Events</th>
<th>Duration</th>
<th>Task-Load</th>
<th>No. of Surge-Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>9</td>
<td>9 min</td>
<td>Moderate</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>12</td>
<td>11 min 45 sec</td>
<td>Moderate</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>18</td>
<td>11 min 51 sec</td>
<td>High</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>11 min 45 sec</td>
<td>Moderate</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
<td>11 min 51 sec</td>
<td>High</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 3-8 was used as a guide to determine event duration within scenarios. For example, an event of priority-level 1 would require at least 70 seconds for its magnitude to reach 0. These minimum times were obtained based on calculations obtained with the CITIES magnitude growth model (A. Wellens & Ergener, 1988). Events programmed into the scenarios for this experiment are required to last longer than the minimum number of time-steps required to complete them in order to accommodate for time required for comprehending event descriptions and completing a response.

**Table 3-8. Minimum time required for event completion**

<table>
<thead>
<tr>
<th>Priority Level</th>
<th>Event Magnitude</th>
<th>Minimum Time-Steps to complete</th>
<th>Minimum Seconds to complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0</td>
<td>23</td>
<td>69 s</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>25</td>
<td>75 s</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
<td>28</td>
<td>84 s</td>
</tr>
</tbody>
</table>

Events within the scenario consist of a description and an event title. Although the term ‘event’ is used here, the participants were provided the term ‘report’ to refer to active events. As mentioned previously, four event categories were present that include, Virus, Worm, Vandalism and Espionage. The specific category definitions (i.e., cues that map to categories), and prioritization rules are provided in the appendix for reference. Category definitions were developed and modified over several iterations before piloting the experiment. The development of event descriptions was initiated after most of the category definitions was fixed, however, the refinement of event descriptions underwent several iterations. A paper pilot session was held within the MINDS Lab (Multidisciplinary Initiatives in Naturalistic Decision Systems) to obtain
feedback from peers on category definitions and priority definitions in training materials and event descriptions. Paper strips containing individual event descriptions without answers were distributed among lab members. Each member provided feedback including a categorization, prioritization and other general feedback. The paper pilot resulted in at least minor changes made to 55% of the events. Table 3-9 includes examples of event descriptions used in the main experiment.

Table 3-9. Examples of events in the scenario

<table>
<thead>
<tr>
<th>Event Title</th>
<th>Event Description</th>
<th>Category</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warehouse Chaos</td>
<td>The warehouse section within the organization uses PCs to keep a record of items flowing in and out. Every system within this department is critical during peak hours. One of the warehouse employees has a problem with a computer that appears to have slowed down significantly over the last 8 hours and it also crashes too often. Additionally, the desktop appears to be filling up with unknown files having a very similar name. The program required for warehouse activity is not usable as its usage is being interrupted with the slowdown and crashing. None of the other systems within the warehouse section appear to be showing similar symptoms so far.</td>
<td>Worm</td>
<td>2</td>
</tr>
<tr>
<td>Nearing a Report Deadline!</td>
<td>An employee in Accounting uses a word processor for writing sections of the company's quarterly report. The employee recently opened a zipped folder containing several images for official use, sent over by an acquaintance through email. Since then the computer would crash every half an hour. Initially, only an internet browser on this workstation was not functional. However, at this point, the word processor is also behaving erratically and is unusable - it would open other documents without requesting them and would close abruptly making it impossible to use.</td>
<td>Virus</td>
<td>3</td>
</tr>
</tbody>
</table>

**Procedure**

The following sequence of steps were followed for running participants for this experiment.

Participants were presented with training slides followed by a training quiz. They were then
presented with answers to a training quiz. A training scenario is provided on the simulation that lasts approximately 9 minutes. Performance scenarios are provided on the simulation where subjective-workload is recorded at the end of each performance scenario. After gathering subjective-workload for the fourth performance scenario, an end-task survey including demographic information, SART (Taylor, 1990), Multitasking Preference Inventory (Poposki & Oswald, 2010) and Big-5 Personality Trait Index (McCrae & John, 1992) are provided in a single survey in the end. The training quiz is a list of 11 multiple-choice questions. These questions were designed towards reducing potential confusion on the category and priority cues, between categories having similar characteristics, whose differences may have been easy to overlook for some participants. Participants are provided with hard-copy slides during the training quiz and during interaction with the simulation for all scenarios so that they do not have to rely on memory for categorizing or prioritizing events in the simulation. The experimental session lasts approximately 2 hours.

**Data Collection and Participants**

The participants were drawn from various Information Sciences and Technology (IST) and Security and Risk Analysis (SRA) undergraduate courses in the College of Information Sciences and Technology and participants were provided extra credit for their participation. Announcements were made in each class to elicit participation. Latin Squares (Senn, 2002) were used in this study to counter balance ordering effects between participants. Table 3-10 provides the number of participants run in each experimental condition in each of the 8 scenario orders considered in the main experimental run. These cells do not include data points that were dropped from the study. Five participants could not be obtained for each cell in Table 3-10 as more
participants could not be obtained and scheduled after dropping participants from the study. More
details on dropped participants are provided in the next chapter.

Table 3-10. Distribution of participants by scenario ordering and experimental condition

<table>
<thead>
<tr>
<th>Latin Square 1</th>
<th>No Workload Preview</th>
<th>Workload Preview</th>
<th>Latin Square 2</th>
<th>No Workload Preview</th>
<th>Workload Preview</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCDA</td>
<td>4</td>
<td>4</td>
<td>ADBC</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DBAC</td>
<td>5</td>
<td>4</td>
<td>CBDA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>CABD</td>
<td>5</td>
<td>5</td>
<td>DACB</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>ADCB</td>
<td>5</td>
<td>5</td>
<td>BCAD</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**Pilot Testing**

A pilot test was conducted with 14 participants prior to the main experimental run. A
significant difference in Normal-Score between Moderate Task-Load and High Task-Load was
obtained where High Task-Load scenarios were associated with lower Normal-Score. No
significant difference in Normal-Score was obtained between scenarios designed for the same
level of Task-Load. Significant differences in Subjective-Workload measured in TLX-Scores
were obtained between Moderate and High Task-Load scenarios where High Task-Load
scenarios received higher TLX-Scores. TLX-Scores for scenarios B and D designed for High
Task-Load were not significantly different, however, Scenario C designed for moderate Task-
Load received a TLX-Score that was significantly lower than that of Scenario A. While these
results indicate that the Task-Load manipulation is effective, it also resulted in modifications
made to scenario A. Events in scenario A were modified with the objective of simplifying
scenario A. A manipulation check item was introduced in the end-task survey to identify task-
switching strategy used by the participant. The training material was modified to add additional
information within the prescribed task-prioritization strategy. The pilot study also encouraged the
use of Normal-Score as the measure of Task-Performance as Raw-Scores (also obtained from the
simulation) did not vary with Task-Load.
Chapter 4

Experimental Results

From a total of 90 participants, 13 participants were removed from the study. Seven participants appeared to have rushed through surveys in the experiment and had to be removed. Three of the participants were outliers in the Normal-Scores – all three had very low Normal-Scores in the training scenario or on one of the performance scenarios. Additionally, three participants were outliers in TLX-Scores – their TLX-Scores were either too high or too low such that no other data-points with comparable TLX-Scores appeared in the sample. After removing these participants from the sample, the sample size is 77, out of which 58 are male. Participants are aged between 18 and 27 years (M = 20.25, SE = .1675). Table 4-1 provides the distribution of the sample by academic year, and, Table 4-2 provides a distribution of the sample by IST, SRA, Computer Science and Engineering (CSE), and other majors.

Table 4-1. Distribution of participants by academic year

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Percentage in Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>19%</td>
</tr>
<tr>
<td>Sophomore</td>
<td>29%</td>
</tr>
<tr>
<td>Junior</td>
<td>29%</td>
</tr>
<tr>
<td>Senior</td>
<td>23%</td>
</tr>
</tbody>
</table>

Table 4-2. Distribution of participants by IST, SRA, CSE and other majors

<table>
<thead>
<tr>
<th>Academic Major</th>
<th>Percentage in Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>IST</td>
<td>44%</td>
</tr>
<tr>
<td>SRA</td>
<td>16%</td>
</tr>
<tr>
<td>CSE</td>
<td>4%</td>
</tr>
<tr>
<td>Other</td>
<td>36%</td>
</tr>
</tbody>
</table>
Analysis of Main Dependent Variables

This section provides an analysis of the main dependent-variables to test the hypotheses stated in Chapter 2. Before proceeding to analyses of main dependent-variables, an estimation of required sample-size is provided in the following sub-section followed by results on Levene’s Test for Equality of Variances.

Estimation of Required Sample-Size

G*Power, a tool to estimate sample size and power was utilized. A desired effect-size of, Cohen’s $d = .8$, was chosen. The corresponding Effect size $f$ value is $.4$, which was provided as input. The experimental design provided as input is - repeated measures ANOVA, with between factors for 4 measurements and 2 groups. An alpha of .05 and power of .95 was chosen as an input. Additionally, a value of $.7$ was chosen for correlation among repeated measures, based on experimental results. The output parameters obtained were, sample-size $= 66$, Critical $F = 3.99$ and Noncentrality parameter $\lambda = 13.62$.

Levene’s Test for Equality of Variances

The Levene’s test for equality of variances indicated significant differences in variance between groups for Task-Prioritization Errors obtained for Scenario A ($F(1,75) = 4.53$, $p = .036$). Variances between the two groups were different because Task-Prioritization Error observations in the Workload-Preview condition had a broader range of values in comparison to the other experimental condition. Analyses on Task-Prioritization Error was conducted with Non-Parametric Tests that do not make the assumption of equal variances across groups. The
following sub-sections provide results on Task-Performance and Task-Prioritization Error. This is followed by a sub-section that will summarize implications of these results to the hypotheses.

**Analysis of Task-Performance**

The Normal-Score provides a measure of Task-Performance. It is calculated by taking an average of Event Normal-Scores on all events for a given scenario. The Event Normal-Score is a function of categorization accuracy, response-time, and priority-level selection accuracy (i.e., selection of levels 1, 2 and 3). Table 4-3 provides the Normal-Scores obtained for each performance scenario. While it is theoretically possible to obtain a Normal-Score approaching a value of 100 (e.g., 99.1), this can very difficult to obtain, and the maximum Normal-Score obtained over every participant and scenario was 67.5 in this experiment.

Scenarios B and D are designed for High Task-Load and Scenarios A and B are designed for Moderate Task-Load. Scenarios C and D are Surge Scenarios whereas Scenarios A and B are Non-Surge scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Normal-Score</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>41.02</td>
<td>1.16</td>
</tr>
<tr>
<td>Scenario B</td>
<td>25.86</td>
<td>1.21</td>
</tr>
<tr>
<td>Scenario C</td>
<td>44.25</td>
<td>1.28</td>
</tr>
<tr>
<td>Scenario D</td>
<td>24.5</td>
<td>1.06</td>
</tr>
</tbody>
</table>

A Shapiro-Wilk test conducted on Normal-Scores obtained for each scenario. It revealed that Normal-Scores are normally distributed. A 2 X 2 X 2 mixed factorial repeated measures analysis of variance (Preview-Availability X Task-Load X Surge-Presence) was conducted to examine Normal-Scores associated with the between (i.e. Preview-Availability) and within subjects (i.e. Task-Load and Surge-Presence) factors. A main effect was observed for Task-Load
where the Normal-Score for Moderate Task-Load (M = 42.64, SE = 1.01) was higher than Normal-Score for High Task-Load (M = 25.18, SE = .97), F (1,75) = 479.6, p < .001, partial $\eta^2 = .86$. A main effect was not observed for Surge-Presence, F (1,75) = 2.62, p = .10, partial $\eta^2 = .034$.

However, a Task-Load X Surge-Presence interaction effect was obtained, F(1,75) = 4.00, p < .05, partial $\eta^2 = .051$. Using Holm’s sequential Bonferroni post hoc comparison, it was found that:

- Normal-Score in Scenario A (M = 41.02, SE = 1.16) was significantly higher than Normal-Score in Scenario B (M = 25.86, SE = 1.21), F (1,75) = 103.66, p < .001, partial $\eta^2 = .58$.
- Normal-Score in Scenario C (M = 44.25, SE = 1.28) was significantly higher than Normal-Score in Scenario D (M = 24.50, SE = 1.06), F (1,75) = 230.99, p < .001, partial $\eta^2 = .75$.
- Normal-Score in Scenario C (M = 44.25, SE = 1.28) was found to be significantly higher than that of Scenario A (M = 41.02, SE = 1.16), F (1,75) = 5.56, p < .05, partial $\eta^2 = .069$. This is an unexpected result.
- Scenarios designed for High Task-Load (i.e., B and D) were not found to be significantly different in Normal-Scores, F (1,75) = 1.30, p = .25, partial $\eta^2 = .017$.

There was no main effect of Preview-Availability on Normal-Score, F(1,75) = .120, p = .730, partial $\eta^2 = .002$. No interaction effects were observed between Preview-Availability, and the within-subjects independent variables. The Workload-Preview did not contribute to improved performance in terms of Normal-Score.
A separate repeated measures analysis of variance was conducted to examine differences in Normal-Score between each scenario. Table 4-4 provides pairwise comparisons for each pair within the 4 performance scenarios.

Table 4-4. Summarization of Normal-Score differences between performance scenarios

<table>
<thead>
<tr>
<th>Scenario A</th>
<th>Scenario B</th>
<th>Scenario C</th>
<th>Scenario D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&gt;B, p &lt; .001</td>
<td>A&gt;C, p = .021</td>
<td>A&gt;D, p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Scenario B</td>
<td>B&lt;C, p &lt; .001</td>
<td>Not Significant</td>
<td></td>
</tr>
<tr>
<td>Scenario C</td>
<td></td>
<td>C&gt;D, p &lt; .001</td>
<td></td>
</tr>
</tbody>
</table>

Overall, Normal-Score was found to be higher under Moderate Task-Load than under High Task-Load. This result does indicate that the Task-Load influenced the main Task-Performance measure. Although, scenarios designed for the same amount of Task-Load were expected to have same performance levels, Scenario C was associated with a higher Normal-Score than Scenario A. Although Task-Performance was hypothesized to be improved with the presence of the Workload-Preview, this hypothesis is not supported.

Analysis of Task-Prioritization Error

Task-Prioritization Errors were obtained for each scenario. The maximum Task-Prioritization Error for scenarios A, B, C and D are eight, twelve, eight and twelve respectively. Table 4-5 provides the means of Task-Prioritization Error for each scenario. Figure 4-1 provides the frequency distribution of Task-Prioritization Error within each scenario. A Shapiro-Wilk test
conducted on Task-Prioritization Error revealed that the distributions significantly deviate from normality (p<.05 for each scenario). Skewness values obtained for Task-Prioritization Error distributions for scenarios A,B,C and D are .738, .339, .385, and .036 respectively. As all are positive skews, it appears that participants were able to prioritize events correctly without significant difficulty.

Table 4-5. Means of Task-Prioritization Error within each scenario

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Task-Prioritization Error</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>1.58</td>
<td>.166</td>
</tr>
<tr>
<td>Scenario B</td>
<td>4.29</td>
<td>.342</td>
</tr>
<tr>
<td>Scenario C</td>
<td>2.14</td>
<td>.179</td>
</tr>
<tr>
<td>Scenario D</td>
<td>4.53</td>
<td>.259</td>
</tr>
</tbody>
</table>

As Task-Prioritization Error’s in each scenario were not normally distributed, a Mann-Whitney test was conducted to test if groups (i.e., Presence vs. Absence of Workload Preview) significantly differed in Task-Prioritization Error. The effect of Workload-Preview was not found to be significant for any of the 4 scenarios. A Wilcoxon Signed Ranks Test (analog to Repeated measures ANOVA) was conducted to examine if Task-Prioritization Error differed with Surge-Presence (a within subjects comparison). For Scenarios A and C, the Wilcoxon Signed Ranks Statistic converted to a Z-score is equal to -2.305. A significant difference was obtained (p<.05). For Scenarios B and Scenario D, the Wilcoxon Signed Ranks Statistic converted to a Z-score is equal to .478. This difference was not found to be significant. Therefore, Task-Prioritization Error under Moderate Task-Load was higher with the presence of surge-events however, a difference in Task-Prioritization Error was not found with the presence of surge-events under High Task-Load.
Hypotheses Testing

Based on the results obtained in previous sub-sections, the following conclusions can be made on the hypotheses.

**H1:** The Workload-Preview will improve Task-Performance.

**Result:** This Hypothesis is not supported. Workload-Preview did not have a main effect on Task-Performance.

**H2:** Task-Performance improvement resulting from the Workload-Preview will be more pronounced in the High-Workload condition.
Result: This Hypothesis is not supported as there was no interaction effect between Workload-Preview and Task-Performance.

H3: The Workload-Preview will improve Task-Prioritization.

Result: This hypothesis was not supported as Workload-Preview did not have a main effect on Task-Prioritization Error on any of the scenarios.

H4: Task-Prioritization improvement resulting from the Workload-Preview will be more pronounced in the High-Workload condition.

Result: Based on the results obtained to test H3 it can be concluded that H4 was not supported.

Clearly, none of the hypotheses stated have been supported. The next chapter discusses this further in more detail. The following sub-section provides experimental results on ancillary experimental variables such as Subjective-Workload and Surge-Prioritization Errors that may help provide resolution on the hypotheses that could not be supported.

Analysis of Ancillary Experimental Variables

Experimental variables not mentioned in the hypotheses are considered to be ancillary. Results obtained on measures incorporated in the experiment such as TLX-Score, Surge-Prioritization Error would provide an insight on why the stated hypotheses are not supported. Moreover, specific results on variables that directly pertain to surge-events would be necessary to analyze specific effects of the Workload-Preview.
Analysis of TLX-Score

The TLX-Score is the measure of Subjective-Workload incorporated in this study. TLX-Score is calculated by taking the sum of 5 scales (also used in NASA-TLX). Table 4-6 provides the means of TLX-Score associated with each performance scenario.

Table 4-6. Means of TLX-Score within each performance scenario

<table>
<thead>
<tr>
<th>Scenario</th>
<th>TLX-Score</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>42.40</td>
<td>1.87</td>
</tr>
<tr>
<td>Scenario B</td>
<td>49.62</td>
<td>1.87</td>
</tr>
<tr>
<td>Scenario C</td>
<td>40.79</td>
<td>1.68</td>
</tr>
<tr>
<td>Scenario D</td>
<td>52.76</td>
<td>2.06</td>
</tr>
</tbody>
</table>

A Shapiro-Wilk test conducted on TLX-Scores obtained for each of the scenarios revealed that TLX-Scores for Scenarios B and D are not normally distributed. Nevertheless, absolute values of skewness calculated for TLX-Scores for each of the scenarios are less than 1.0. TLX-Scores obtained for scenarios A, B, C and D have skewness values of -.255, -.371, -.134, -.344, respectively. As groups by experimental condition were nearly equal in size, analyses of variance that may be robust to deviations from normality are reported for TLX-Scores.

A 2 X 2 X 2 mixed factorial repeated measures analysis of variance was conducted to examine the effect of Preview-availability, Task-Load and Surge-Presence on TLX-Score. A main effect was observed for Task-Load, where the Moderate Task-Load group (M = 41.59, SE = 1.66) was associated with lower TLX-Score than the High Task-Load group (M = 51.19, SE = 1.84), F(1,75) = 84.51, p < .001, partial η² = .53. A main effect was not observed for Surge-Presence, F(1,75) = 1.07, p = .303, partial η² = .014. However, a Task-Load X Surge-Presence interaction effect was obtained, F(1,75) = 4.56, p < .05, partial η² = .057. Using Holm’s sequential Bonferroni post hoc comparison, it was found that:
• TLX-Score associated with Scenario B (M = 49.62, SE = 1.87) was significantly higher than that of Scenario A (M = 42.40, SE = 1.87), F(1,75) = 22.55, p<.001, partial $\eta^2 = .231$.

• TLX-Score associated with Scenario D (M = 52.76, SE = 2.06) was significantly higher than that of Scenario C (M = 40.79, SE = 1.68), F(1,75) = 61.55, p < .001, partial $\eta^2 = .451$.

• Scenario A and C did not significantly differ in TLX-Score, F(1,75) = 1.59, p = .210, partial $\eta^2 = .021$.

• TLX-Score associated with Scenario B (M = 49.62, SE = 1.87) was significantly lower than that of Scenario D (M = 52.76, SE = 2.06), F (1,75) = 5.08, p = .027, partial $\eta^2 = .063$.

No main effect was observed for Preview-Availability on TLX-Score, F (1,75) = .143, p=.706, partial $\eta^2 = .002$. However, a Task-Load X Preview-Availability interaction effect is approaching significance, F (1,75) = 3.55, p = .063, partial $\eta^2 = .045$.

A separate repeated measures analysis of variance was conducted to examine differences in TLX-Score between each scenario. Table 4-7 provides pairwise comparisons for each pair within the 4 performance scenarios.

Table 4-7. A summarization of TLX-Score differences between scenarios

<table>
<thead>
<tr>
<th>Scenario A</th>
<th>Scenario B</th>
<th>Scenario C</th>
<th>Scenario D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>A&lt;B, p&lt;.001</td>
<td>Not Significant</td>
<td>A&lt;D, p&lt;.001</td>
</tr>
<tr>
<td>Scenario B</td>
<td>B&lt;C, p&lt;.001</td>
<td>B&lt;D, p = .027</td>
<td></td>
</tr>
<tr>
<td>Scenario C</td>
<td></td>
<td>C&lt;D, p&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>
Analysis of Surge-Prioritization Error

Scenario’s C and D consisted of secondary task events that increased in severity immediately after initiation. Their severity exceeds that of concurrent primary tasks. These events are referred to as surge-events and prioritization errors were obtained based on the response sequence to surge-events and other concurrent events. The Surge-Prioritization Error is similar to the Task-Prioritization Error but is specific to surge-events for a given scenario.

Table 4-8 provides the Mean values of Surge-Prioritization Error obtained for each scenario consisting of surge-events. According to Shapiro-Wilk test, Surge-Prioritization Error on Scenario’s C and D were not normally distributed (p<.05 for each). Skewness values obtained for Surge-Prioritization Error for Scenario’s C and D were, -.763 and -.326. Figure 4-2 provides frequency-distributions of Surge-Prioritization Error obtained for scenarios C and D.

Table 4-8. Mean Surge-Prioritization Score within each condition

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Surge-Prioritization Error</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario C (Moderate Task-Load)</td>
<td>1.32</td>
<td>.143</td>
</tr>
<tr>
<td>Scenario D (High Task-Load)</td>
<td>2.66</td>
<td>.2</td>
</tr>
</tbody>
</table>
As Surge-Prioritization Error frequency-distributions for the two scenarios significantly deviate from normality, for scenarios C and D, each, a Mann-Whitney U test was conducted to test if the groups (i.e. with and without Workload-Preview) significantly differed in (medians of) Surge-Prioritization Error. For Scenario C, The Mann-Whitney U statistic is equal to 734, and $p = .941$. Therefore, for scenario C, the two groups did not significantly differ in Surge-Prioritization Error.

For Scenario D, The Mann-Whitney U statistic is equal to 500.5, and $p<.05$. Therefore, for scenario D, the two groups (i.e., presence vs. absence of Workload-Preview) significantly differed in Surge-Prioritization Error. However, the direction of the difference is opposite of the anticipated effect of the Workload-Preview. The group with the Workload-Preview committed significantly more Surge-Prioritization Errors than the group without it. This result indicates that the Workload-Preview may have been detrimental to the ability to take notice of secondary task surges.
Analysis of Surge Normal-Score

For each scenario consisting of secondary task surges (i.e., Scenarios C and D), the mean of Event Normal-Scores for secondary task surge-events were calculated. This variable is referred to as Surge Normal-Score.

The Levene’s test of equality of variances indicated significant differences in variances between the two experimental conditions for the Surge Normal-Score in Scenario C (F(1,75) = 7.79, p < .01). The variance in the Workload-Preview condition was higher as observations in this condition had a broader range of values. Surge Normal-Score in Scenario C is not normally distributed according to the Shapiro-Wilk Test (p = .001), and its skewness value is -.982. However, Surge Normal-Score in Scenario D is normally distributed and it is associated with a skewness value of .219 (below 1.0). Figure 4-3 provides frequency-distributions of Surge Normal-Score for Scenarios C and D. Although, it is desired to have a normal distribution, and equal error variances across groups, analyses of variance can be robust to this if group sizes are nearly equal.

Figure 4-3. Frequency Distributions for Surge Normal-Scores for scenarios C and D
A 2X2 mixed factorial repeated measures ANOVA was conducted to examine the effect of Preview-Availability and Task-Load on Surge Normal-Score. Although the condition with the Workload-Preview (M = 33.42, SE = 1.81) received a lower Surge Normal-Score than the condition without the Workload-Preview (M = 37.42, SE = 1.78), this difference was not significant, $F(1, 75) = 2.47, p = .120$, partial $\eta^2 = .032$. There was no interaction effect obtained between Task-Load and Preview-Availability, $F(1, 75) = .001, p = .97$, partial $\eta^2 = 0.0$. This result indicates that the Workload-Preview did not have a positive effect on Task-Performance on surge-events. Instead, the results suggest that the Workload-Preview may have negatively affected Task-Performance on surge-events.

**Predictors of Task-Performance**

Testing for relationships that exist between Task-Performance and other variables may provide insight about factors that determined Task-Performance. Table 4-9 provides the results of a Linear Regression analysis that tests for the strength of, mental-effort as a predictor of Normal-Score. Table 4-10 provides the strength of Surge-Prioritization Error as a predictor of Surge-Normal Score.

Table 4-9. Linear Regression between Normal-Score (DV) and mental-effort (predictor)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Linear Regression Statistic</th>
<th>Beta Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td>$F(1, 75) = 1.95, p = .166$</td>
<td>-.160</td>
</tr>
<tr>
<td>Scenario B</td>
<td>$F(1, 75) = 4.07, p = .047$</td>
<td>-.227</td>
</tr>
<tr>
<td>Scenario C</td>
<td>$F(1, 75) = 4.39, p = .039$</td>
<td>-.235</td>
</tr>
<tr>
<td>Scenario D</td>
<td>$F(1, 75) = 1.85, p = .177$</td>
<td>-.155</td>
</tr>
</tbody>
</table>
Table 4-10. Linear Regression between Surge Normal-Score (DV) and Surge-Prioritization Error (predictor)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Linear Regression Statistic</th>
<th>Beta Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario C</td>
<td>$F (1, 75) = 34.00, p &lt; .001$</td>
<td>-.559</td>
</tr>
<tr>
<td>Scenario D</td>
<td>$F (1, 75) = 40.91, p &lt; .001$</td>
<td>-.594</td>
</tr>
</tbody>
</table>

Table 4-9 indicates that mental effort predicted Task-Performance in Scenarios B and C. This relationship was not detected in Scenarios A and D as the Linear Regression for mental-effort as a predictor of Task-Performance, was not found to be significant.

Table 4-10 indicates that Surge-Prioritization Error is a strong predictor of Surge Normal-Score. Individuals with fewer Surge-Prioritization Errors tended to perform better on surge-events in terms of reaction-time and event categorization accuracy.
Summarization of Results

Table 4-11 below summarizes the analysis reported in this chapter on main and ancillary dependent variables.

Table 4-11. Summarization of Experimental Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Key Results from Analysis</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-Score</td>
<td>• Did not differ with Workload-Preview</td>
<td>H1 and H2 is not supported</td>
</tr>
<tr>
<td></td>
<td>• Significantly differs with Task-Load</td>
<td></td>
</tr>
<tr>
<td>Task-Prioritization Error</td>
<td>• Did not differ with Workload-Preview</td>
<td>H3 and H4 is not supported</td>
</tr>
<tr>
<td></td>
<td>• Significantly differs with Surge-Presence</td>
<td></td>
</tr>
<tr>
<td>TLX-Score</td>
<td>• Subjective-Workload significantly differs with Task-Load</td>
<td>High Task-Load results in increased mental-workload</td>
</tr>
<tr>
<td></td>
<td>• No main effect of Workload-Preview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Near-significant interaction effect between Workload-Preview and Task-Load</td>
<td></td>
</tr>
<tr>
<td>Surge-Prioritization Error</td>
<td>Significantly differs with Workload-Preview on Scenario D - Surge</td>
<td>The Workload-Preview results in degraded performance in the prioritization of surge-events. This is the opposite of what was expected based on the hypotheses.</td>
</tr>
<tr>
<td></td>
<td>Prioritization was lower in the Workload-Preview condition</td>
<td></td>
</tr>
<tr>
<td>Surge-Normal Score</td>
<td>Workload-Preview did not have a main effect on Surge-Normal Score</td>
<td>Workload-Preview did not contribute to improved Task-Performance on surge-events. This is the opposite of what was expected based on the hypotheses.</td>
</tr>
</tbody>
</table>
Chapter 5

Discussion

The experiment was designed to create unexpected variations in secondary-task priority indicated to the participant in the form of Overall-Severity. This is a problem characteristic in a number of domains including network monitoring in cyber-security. The experimental study manipulates task-load and the presence of surge-events in scenarios. The effect of a Workload-Preview is examined on various measures of performance within this context. This chapter provides an interpretation of experimental results. Questions including, were secondary-task surge-events unexpected in scenarios C and D?, and, was the ability to expect surge-events in secondary-task events affected by the Workload-Preview or Task-Load?, are addressed in this chapter. Towards the end of the chapter, a resolution on the hypotheses is provided.

Task-Prioritization and Surge-Event Effects

Task-Prioritization Error is the number of events that were responded to during the presence of a higher priority concurrent event. The number of prioritization errors on surge-events specifically is referred to as Surge-Prioritization Errors. Frequency distributions of Task-Prioritization Error in scenarios A, B, C and D have skewness values of .738, .339, .385 and .036 respectively. These positive skews and the frequency distributions provided in Figure 4-1, indicate that the participants could maintain the prescribed task-prioritization strategy in scenarios A, B, and C easily. Further, frequency distributions (i.e., Figure 4-2) of Surge-Prioritization Error in scenarios C and D, and, their skewness values of .763 and .326, also indicate that maintaining prescribed prioritization strategy for surge-events was easy for a large number of participants. While these results indicate that participants understood and applied the material presented in the
training slides, an objective underlying the experimental design was to make surge-prioritization difficult - this difficulty should have been associated to not allowing the participant to expect the surge-events easily. While Surge-Prioritization Error only takes into account the sequence in which events were responded to, Surge Normal-Score takes response-time, accuracy, and priority-level assignment into account as well. Figure 4-3 in the previous chapter provides a frequency-distribution of Surge-Prioritization Error. It appears that surge-events were easy to identify, prioritize and answer in a timely manner, although relatively less easily in scenario D.

Task-Prioritization Error in Scenario C (M = 2.14, SE = .179) was significantly higher than that of Scenario A (M= 1.58, SE = .166) based on a Wilcoxon signed ranks test (Z = -2.30, p < .05). The maximum number of Task-Prioritization Errors that can be received for either scenario is 8. To test for significant differences in subjective ratings of workload between scenarios A and C, a repeated measures analysis of variance was conducted with the 5 NASA-T LX dimensions (excluding performance) as dependent-variables and Preview-Availability as a covariate – subjective-workload ratings along any of the dimensions are not significantly different. The difference in Task-Prioritization Error obtained between scenarios A and C can be attributable to the presence of surge-events, although, no significant difference on Task-Prioritization Error was obtained between scenarios B and D, based on a Wilcoxon signed ranks test.

A 2X2X2 mixed factorial analysis of variance indicated a significant difference in TLX-Score between scenarios B and D, although the difference is associated with an effect-size much smaller (.063) than the effect-sizes obtained upon comparing scenarios of different task-load (.231, .451). To examine the specific dimensions that were rated differently between scenarios B and D, a mixed factorial analysis of variance was conducted with the 5 Subjective-Workload dimensions as dependent variables and Preview-Availability as a covariate. Mental-effort in Scenario D was rated to be higher than mental-effort in Scenario B and this difference is
approaching significance (F(1,75) = 3.07, p = .084, partial η² = .039). None of the ratings in the Subjective-Workload instrument were approaching significance. Surge-Events in scenario D constitute 3 out of the 18 events. The presence of surge-events appears to have required increased mental-effort within Scenario D.

To summarize, the presence of surge-events in scenario C was associated to more prioritization errors in comparison to those in scenario A. Although Scenarios B and D did not differ in the number of prioritization errors, the presence of surge-events in scenario D appears to have required more mental-effort. Given that one of the surge-events in scenario D appeared at the beginning of the scenario within the first batch of events, it is possible that participants expected more surge-events later in the scenario resulting in increased visual scanning of the Task-State Overview. This may be the cause of increased mental-effort, however, the difference in mental-effort between scenarios B and D approaches significance and does not have a large effect size (partial η² = .039). Overall, it appears that surge-events in Scenarios C and D were predictable by the participants and they may have been aware of when to scan for potential surges in the secondary task.

**Effect of Task-Load**

Scenarios A and C were designed for Moderate Task-Load, and Scenarios B and D were designed for High Task-Load. The difference in the two levels of Task-Load was created by compressing more events in the High Task-Load scenarios. An analysis of the TLX-Scores indicates that the higher Task-Load scenarios were subjectively assessed to involve higher Subjective-Workload. To assess which dimensions contributed to higher TLX-Scores, another repeated measures analysis of variance was conducted with all Subjective-Workload ratings as dependent-variables and Task-Load as the within subjects variable – the comparison between
Scenario A and B indicates that all subjective ratings were significantly different \( (p<.01) \) except for subjective estimate of mental-effort; however, temporal-demand (partial \( \eta^2 = .187 \)), frustration (partial \( \eta^2 = .171 \)) and mental-demand (partial \( \eta^2 = .151 \)) had the highest effect sizes. Similarly, the comparison between scenario C and D indicates that all subjective ratings were significantly different \( (p<.01) \); however, temporal-demand (partial \( \eta^2 = .498 \)) and mental-demand (partial \( \eta^2 = .387 \)) had the highest effect sizes.

Overall, the analysis of ratings obtained with the Subjective-Workload instrument indicates that higher Task-Load was associated to increased temporal-demand, mental-demand and even frustration.

**Effect of Workload-Preview**

The Workload-Preview was anticipated to provide a larger window of opportunity for being able to notice changes in secondary-task severity. The hypotheses state that the presence of the Workload-Preview would result in lower Task-Prioritization Error and improved Task-Performance. Moreover, it was anticipated that this effect would be more pronounced within surge-events. The experimental results provided in the previous chapter indicate that the Workload-Preview did not result in any of these hypothesized improvements (see page no. 79). The presence of the Workload-Preview was anticipated to reduce mental-workload resulting from reduced visual scanning of the Task-State Overview and related information integration. However, there was no main effect of Workload-Preview on TLX-Score.

The (main) effect of Preview-Availability was not significant. This indicates that the Workload-Preview did not mitigate mental-workload; however, on TLX-Score (i.e., subjective-
workload) an interaction between Preview-Availability and Task-Load was found to be
approaching significance ($F(1,75) = 3.55, p = .063, \text{partial } \eta^2 = .045$).

Figure 5-1 below describes this interaction effect. However, post-hoc comparisons did
not reveal (significant) effects of Preview-Availability on TLX-Score within individual levels of
Task-Load.

Based on this interaction effect, it appears that there is a shift on the effect of the
Workload-Preview on mental-workload from Moderate to High Task-Load. Upon increasing
Task-Load, the Workload-Preview appears to contribute to increased mental-demand relative to
the non Workload-Preview condition. However, the observed increase in TLX-Score in the High
Task-Load condition is not significant. Specifically, the Workload-Preview possibly results in
increased mental-demand with increased Task-Load ($F(1,75) = 2.89, p = .098, \text{partial } \eta^2 = .036$)
although, mental-effort does not increase to the same extent with the use of the Workload-Preview with increased Task-Load (F(1,75) = 1.45, p = .232). The Workload-Preview did not facilitate overall Task-Performance measured by Normal-Score. On Surge Normal-Scores specifically, the group without the Workload-Preview received a higher mean Surge Normal-Score than the group with the Workload-Preview. However, this difference does not approach significance (p = .12). The Workload-Preview did not result in reduced Task-Prioritization Error overall. Moreover, results on Surge-Prioritization Error are opposite of what was anticipated. On Scenario D, differences on Surge-Prioritization Error between the two conditions indicate that the Workload-Preview may have been detrimental to the prioritization of surge-events - individuals provided with the Workload-Preview made significantly more Surge-Prioritization Errors than individuals without it. Figure 5-2 provides the Surge-Prioritization Error frequency-distribution within each experimental condition. The group without the Workload-Preview appears to have performed well in the prioritization of surge-events - the frequency-distribution has a positive skewness of .508. The group with the Workload-Preview has a skewness of .025 and observations in Surge-Prioritization Error in the Workload-Preview group are more spread-out. A non-parametric test indicates that the effect of the Workload-Preview is significant. The Medians of Surge-Prioritization Errors in Scenario D are 2 and 3 in the non Workload-Preview condition and Workload-Preview condition, respectively. This suggests that the Workload-Preview was detrimental to the prioritization of surge-events in Scenario D.
Based on analyses on Subjective-Workload, there is a shift on the effect of the Workload-Preview on TLX-Score. An effect of Workload-Preview is only seen on Surge-Prioritization Error and not on Task-Prioritization Error. It appears that the Workload-Preview contributed to increased mental-demand that resulted in reduced scanning of the Task-State Overview resulting in missing the oncoming surge-events. Participants followed a fixed or ‘blind’ strategy of giving second priority to the secondary-task events that failed when surge-events were present, however, this fixed or ‘blind’ prioritization strategy avoided increases in Task-Prioritization Error in Scenario D or Scenario B.

The difference between the experimental conditions in Surge-Prioritization Error in Scenario C was not significant. In Scenario C, participants in the Workload-Preview condition may have used the Task-State Overview, with or without the Workload-Preview component, to prioritize events in the scenario. As Scenario C was a Moderate Task-Load scenario, there may have been enough mental resources to process information from the Workload-Preview. As Surge-Prioritization under Moderate Task-Load was a relatively easy sub-task, the use of the Workload-Preview in Scenario C did not contribute to improved performance in Surge-Prioritization in comparison to the non Workload-Preview condition.
However, a High Task-Load scenario, involving increased mental-demand and temporal-demand (see page no. 91) may have prevented the use of the Task-State Overview as the Workload-Preview may have added to the perceived complexity of the Task-State Overview. To summarize, the relative complexity of the Task-State Overview, with the Workload-Preview, resulted in reluctance of its usage thereby causing participants to follow a fixed prioritization strategy. This strategy possibly resulted in increased errors in surge prioritization without affecting the prioritization of regular events.

**Effects of Learning**

In the experiment, there were eight possible orderings for the four performance scenarios. Although, a training scenario was provided before the four performance scenarios, effects of learning may influence performance measures. This between-subjects analysis compares performance measures taken from scenarios placed at different positions in the scenario order sequence.

Based on Figure 5-3, Task-Performance within each scenario, appears to show an effect of learning. Shapiro-Wilk tests were conducted, and Normal-Score for Scenario C was not normally distributed at the first scenario position (p < .01). The cause of deviation from Normality is the presence of outliers within that group. As group sizes are relatively equal and ANOVA can be robust to deviations from Normality, Univariate analyses were conducted to test for effects of learning. All other distributions of Normal-Score within each group are normal. For each performance scenario, a Univariate analysis of variance was conducted to examine the effect of scenario position on Normal-Scores. For each scenario, effects of position on Normal-Score (between-subjects) were significant. The partial $\eta^2$ values obtained were .327, .194, .308, and .114 for scenarios A, B, C and D respectively; and p<.05 for each scenario.
On the other hand, an effect of learning in Task-Prioritization Error was not consistently present across all scenarios. Figure 5-4 shows mean Task-Prioritization Error obtained at different scenario positions for each scenario. At a glance, the four performance scenarios do not appear to share a common trend. For each scenario, a Kruskal-Wallis test was conducted to examine the effect of scenario position on Task-Prioritization Error. This non-parametric test was chosen because Task-Prioritization Error is not normally distributed. No main effect of scenario position is revealed for Scenario A (Chi-Square = 2.95, p = .398). A main effect of scenario position is present in Scenario B (Chi-Square = 14.83, p = .002). No main effect of scenario position was revealed for Scenario C (Chi-Square = .085, p = .994), and, no main effect of scenario position was revealed for Scenario D (Chi-Square = 1.21, p = .75). In the case of Scenario B, Task-Prioritization Errors between first and second positions (an unexpected difference seen in Figure 5-4) were not significantly different based on a Mann-Whitney test (Z = -1.24, one-tailed p = .221); however, Task-Prioritization Error between second and fourth positions are significantly different based on a Mann-Whitney test (Z = -2.99, one-tailed p = .003).

![Figure 5-3. Mean Normal-Score across scenario position (between-subjects)](image-url)
Figure 5-4. Mean Task-Prioritization Errors across scenario position (between-subjects)

Figure 5-5 and Figure 5-6 describe relative differences in performance measures obtained for surge-events, between scenario positions. All Latin-Square orderings were such that Scenarios C and D were never both in the first half or both in second half of scenario positions.

Figure 5-5. Mean Surge Normal-Score trend across scenario position (between-subjects)
A one-way analysis of variance was conducted to examine the effect of Scenario C’s position on Surge Normal-Score for Scenario C. The result of this test revealed that Scenario C’s position had a main effect on Surge Normal-Score for Scenario C (F(3, 73) = 4.46, p = .006, partial $\eta^2 = .155$). Pairwise comparisons indicate that Surge Normal-Scores for Scenario C significantly differed between first and fourth positions (p < .01), and, first and third positions (p = .046). Surge Normal-Scores for Scenario C did not differ significantly between first and second positions. The significant effect of learning in Scenario C observed particularly in the latter 2 positions is possibly an effect of Scenario D being performed earlier (refer to Figure 5-5 above).

A one-way analysis of variance was conducted to examine the effect of Scenario D’s position on Surge Normal-Score for Scenario D. Scenario D’s position did not have a main effect on Surge Normal-Score (F(3, 73) = .892, p = .449). This indicates that there is no effect of learning on Surge-Normal Scores in Scenario D.

Figure 5-6. Mean Surge-Prioritization Error trend across scenario position (between-subjects)
A Kruskal-Wallis test was conducted to examine if scenario position for C affected Surge-Prioritization Error. The test revealed that Surge-Prioritization Error does not significantly differ between different positions for Scenario C in their respective median values (Chi-Square = .688, p = .876). A Kruskal-Wallis test was conducted to also examine if scenario position for D affected Surge-Prioritization Error. The test revealed that Surge-Prioritization Error does not significantly differ between scenario positions for Scenario D (Chi-Square = .908, p = .823). Kruskal-Wallis tests were used as Surge-Prioritization Error is not normally distributed.

Table 5-1. Summary on effect of learning

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Summary of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task-Performance (Normal-Score)</td>
<td>An effect of learning is present on each scenario.</td>
</tr>
<tr>
<td>Task-Prioritization Error</td>
<td>Effect of learning is not present on 3 scenarios. An effect of learning is present on Scenario B.</td>
</tr>
<tr>
<td>Surge Normal-Score</td>
<td>An effect of learning is present on Scenario C but not on Scenario D.</td>
</tr>
<tr>
<td>Surge Prioritization Error</td>
<td>Effect of learning not present on Scenarios C and D.</td>
</tr>
</tbody>
</table>

Based on these results, it appears that the prescribed Task-Prioritization strategy was relatively easy to comprehend and apply. However, performance in assigning categories and priorities to events improved with time – it is possible that this effect of learning may be present in either reaction-times or accuracy or both and this has not been analyzed.
Task-Performance and Normal-Score

As stated in the previous chapter, according to Shapiro-Wilk tests, Normal-Scores were normally distributed in each scenario including the training scenario. Figure 5-7 provides the frequency-distributions of Normal-Score obtained for each performance scenario.

![Figure 5-7](image)

Figure 5-7. Normal-Scores obtained on each scenario

The Normal-Score could capture effects of learning in Task-Performance discussed in a previous section (refer to page no. 96). In particular, in Scenario D that had the smallest difference in mean values of Normal-Score, a main effect of scenario position on Normal-Score was observed ($F (3, 73) = 3.13, p = .031, \text{ partial } \eta^2 = .114, \text{ observed power } = .70$). Effects of
learning were not observed for Task-Prioritization Error. Skewness values obtained for Normal-Score distributions in the training-scenario, and, Scenarios A, B, C, and D, are, -.013, -.054, .259, -.041, and .169.

Normal-Score for Scenario C was found to be higher than Normal-Score for Scenario A (F (1,75) = 5.56, p < .05). This was an unexpected difference. A similar effect was observed in the pilot test as well. In the pilot test, scenario C received a TLX-Score that was significantly lower than that of scenario A. For the main experiment, changes to event-descriptions were made within Scenario A to simplify events within Scenario A in order to balance out the two scenarios in terms of difficulty-level. Although, it is possible that events in Scenario C were easier than the ones in Scenario A, an analysis of TLX-Score does not reveal a significant difference.

The Normal-Score based on the magnitude-growth model used in CITIES based simulations was used in different contexts prior to this experiment. In NeoCities, it has been used within the context of resource-allocation and it mainly corresponds to a measure of team-performance. NETS-DART uses this as individual measure of Task-Performance and the concept of resource or unit allocation is replaced by categorization and priority-level assignment. The Normal-Score is a function of response-time, category-assignment accuracy, and priority-level-assignment accuracy. Response-time and priority-level are taken into account upon correct category assignment. The Normal-Score frequency-distributions obtained in this experiment are encouraging and they indicate that these could be used as measures of task-performance in future studies.

Resolution on the Hypotheses

As stated in the hypothesis testing section in the previous chapter, the results indicate that none of the hypotheses regarding the effect of the Workload-Preview stated in Chapter 2 have
been supported. Hypothesis 1 states that Task-Performance would improve with the Workload-Preview and Hypothesis 3 states that Task-Prioritization would improve with the Workload-Preview.

The mental-demands associated with Scenarios A and C were not sufficient for the Workload-Preview to have a potential benefit in terms of Task-Performance or Task-Prioritization. Task-Prioritization including Surge-Prioritization was anticipated to involve difficulty however, the participants were successful in comprehending the task-prioritization strategy instructed during training, and applying it. Perhaps there would have been a benefit of the Workload-Preview in a slightly more complex scenario with unexpected or unpredictable surge-events that may have required frequent scanning of the Task-State Overview.

In Scenario A, mental-demand and temporal-demand were not sufficient enough for the Workload-Preview to have a large benefit in terms of Task-Performance or Task-Prioritization. Additionally, given that Scenario A did not consist of surge-events (although participants are not aware of this condition) it is possible that there was less anticipation among participants about surge-events. The 12 second window of opportunity may have not resulted in reduced scanning of the Task-State Overview to the extent of benefitting the main task of categorizing and answering events. Any Task-Prioritization improvement resulting from the Workload-Preview in Scenario A would have been limited to benefit of having additional time to select the sub-task before each batch of events appeared. As many participants in either experimental condition may have used a strategy of giving automatic preference to the primary task without referring to the Task-State Overview, the effect of the Workload-Preview was negligible.

Similarly, Scenario C did not create the mental-demand or temporal-demand sufficient enough for the Workload-Preview to have a large benefit on Task-Prioritization and Task-Performance. The mean of Task-Prioritization Errors in Scenario C was lower in the Workload-Preview condition (M = 1.47, σ = 1.97) in comparison to the mean of that in the non Workload-
Preview condition (M = 1.65, σ = 2.30), although this difference is not statistically significant. Scenario C consisted of surge-events. Surge-Prioritization Errors in the two conditions within Scenario C were nearly equal – Scenario C was possibly not demanding enough for the additional 12 second window of opportunity to result in freed mental-resources causing an improvement in task-performance and task-prioritization. Task-Performance measured by Normal-Score was also nearly equal between the two experimental groups.

Task-Load differences between scenarios resulted in higher TLX-Scores in Scenarios B and D. Scenarios designed for higher Task-Load involved higher mental-demand and temporal-demand as rated by participants (refer to page no. 91). The Task-Load manipulation was successful in creating the effect that was desired. Increased Task-Load resulted in both degraded Task-Prioritization and Task-Performance and this effect although not explicitly stated in the hypotheses, was anticipated based on previous experimental studies on Task-Management that have manipulated workload (Andre, et al., 1995; Metzger & Parasuraman, 2005; Raby & Wickens, 1994).

Hypothesis 2 stated that the Workload-Preview would result in more pronounced improvement in Task-Prioritization, and Hypothesis 4 stated that the Workload-Preview would result in more pronounced improvement in Task-Performance. In the case of Scenario B (High Task-Load and No Surge-Events), Means of Task-Prioritization Error are nearly equal across the two experimental conditions. Similarly, means of Normal-Score are nearly equal between the two conditions. The prescribed task-prioritization strategy may have been applied without significant use of the Task-State Overview. Even if a potential benefit resulted from the use of Workload-Preview, this benefit would have been small in the case of Scenario B as surge-events were not present.

Scenario D consisted of surge-events. It appears that the Workload-Preview degraded task-prioritization performance on surge-events. Surge-Prioritization Errors are significantly
higher in the Workload-Preview condition. The interaction-effect on mental-workload between Task-Load and the Preview-Availability indicates that preview information may have been associated to higher mental-demand. Consequently, participants withdrew from the use of the Task-State Overview under High Task-Load in Scenario D where participants followed a fixed ‘blind’ task-prioritization strategy that involved giving default (high) priority to the primary-task without comparing severity levels from the Task-State Overview. While this withdrawal involving the fixed ‘blind’ prioritization strategy may have been present in the Workload-Preview condition on Scenario B as well, the difference in Task-Prioritization Error in Scenario B was not significant possibly because surge-events were not present in Scenario B. Therefore, in Scenario D, the withdrawal from scanning the Task-State Overview in Workload-Preview condition may have been associated to increased spare mental capacity that should have been allocated to attending to regular (non-surge) events. This should have resulted in the Workload-Preview group receiving higher Normal-Scores on non-surge-events in comparison to that on the non Workload-Preview group. Normal-Score means were calculated for non-surge-events in Scenario D. In Scenario D, for non-surge-events, the mean Normal-Score was found to be higher in the Workload-Preview group (M = 26.27, \( \sigma = 10.49 \)) in comparison to that of the non Workload-Preview group (M = 23.16, \( \sigma = 9.51 \)), although this difference is not significant according to a one-way analysis of variance (F(1,75) = 1.86, p = .17). Moreover, an opposite trend is observed on Surge Normal-Scores in Scenario D – The Workload-Preview group received a lower Surge Normal-Score (M = 20.90, \( \sigma = 13.93 \)) in comparison to that of the non Workload-Preview group (M = 26.09, \( \sigma = 13.42 \)) although this effect is also not significant (F(1,75) = 2.76, p = .10).

Although these pairwise comparisons on Normal-Score between the Non-Workload-Preview group and the Workload-Preview group in Scenario D did not differ within either types of event, the interaction effect between event-type and Preview-Availability (as in Figure 5-8) is significant, F(1,75) = 4.91, p = .03, partial \( \eta^2 = .061 \). This interaction-effect indicates a ‘shift’ in
attention-allocation strategy between the non-Workload-Preview group and the Workload-
Preview group.

Figure 5-8. Normal-Scores for Regular-Events and Surge-Events in Scenario D across Preview-
Availability

Previous literature on the use of previews has shown that previews that present
information that requires real-time mental processing can be detrimental to task-performance
(Cummings & Mitchell, 2005; Wickens, et al., 1991), although this has not been explored
specifically on task-prioritization performance. The results obtained in this study on surge-
prioritization within the High Task-Load condition indicate that the Workload-Preview
incorporated in this study was detrimental to performance. Normal-Scores obtained on surge-
events also indicate that the Workload-Preview may have been detrimental to task-performance
within surge-events although this result only approaches significance. While this finding is
similar to other findings in the literature, there are several major differences – the Workload-
Preview incorporated here is perfectly reliable and does not involve interpretation of probabilistic
information. The failure of previews on improving performance reported in the literature is attributed to increased cognitive load resulting from probabilistic information or imperfect reliability. Moreover, the Workload-Preview may have negatively affected only one component of the task (although an important one) – the prioritization and response to surge-events. Task-Performance and Task-Prioritization on the overall task was unaffected by the Workload-Preview.
Chapter 6

Conclusion

The stated hypotheses about the potential effects of the Workload-Preview were not supported. Moreover, the Workload-Preview may have degraded task-prioritization according to results in the High Task-Load scenario consisting of surge-events. In this chapter, implications from these results are addressed. Although the hypotheses were not supported, the previous chapter provides a resolution on this (refer to page no. 102). Additionally, several results from this experiment regarding the effect of Task-Load were anticipated based on the literature, although not explicitly stated as hypotheses. This chapter addresses various implications from experimental results of this dissertation. The following sections will address the implications of this study to Human Factors and Cyber Security. This is followed by a discussion on the limitations, and, future work.

Implications to Human-Factors

An objective of this experiment was to explore the effectiveness of a Workload-Preview as a cognitive-aid for attention-guidance. While many attention-guiding aids are successful, the effectiveness of an attention-guiding aid is moderated by contextual factors such as unexpected shifts in priorities and external task-load. The result suggests that the Workload-Preview appears to have degraded the ability to successfully prioritize secondary task events that increased in severity unexpectedly (i.e., surge-events). The previous chapter discusses the possibility that this may have resulted from withdrawal from the use of the Task-State Overview altogether (refer to page no. 102) as the Workload-Preview added to the complexity of the display. With increased cognitive-load resulting from temporal and mental demand from the High Task-Load scenario,
the tendency to withdraw from scanning the Task-State Overview (especially in the Workload-Preview condition) was possibly higher mainly due increased mental-demand resulting from it. The component of the Task-State Overview providing current information on Overall-Severity comparison is sufficient to guide task-prioritization. The preview that provides a larger window of opportunity to notice secondary-task surges was anticipated to make more mental capacity available; it was however, not utilized by a percentage of the Workload-Preview group mainly owing to increased mental-demand associated to cognitive processing requirements associated to its use. An implication from this study to design of attention-guiding aids is that the format should afford mental processing of preview information without additional cognitive-load. Individuals tend to satisfice task goals and prioritization rules. There is less of a tendency to optimize any variable, unless a visible outcome results from such optimization (e.g. a score). In this task, the use of preview information can be considered to be closer to optimization-behavior in comparison to satisfying-behavior, as current Overall-Severities, present in both experimental conditions, was sufficient to determine the requirement to switch tasks and accordingly follow the prescribed prioritization strategy. It appears that withdrawal from information that tends to optimize task-performance is coupled with withdrawal from related elements in the display in close proximity to it, as individuals mentally integrate related elements together and view them as one object (Wickens & Hollands, 2000). An implication of this result for the design of level-1 (Parasuraman, Sheridan, & Wickens, 2000) cognitive-aids is that information that primarily facilitates optimization of performance can be separated from elements on the display that are critical for satisfying specific task-goals or criterion. Operators can choose to utilize information intended to optimize performance under moderate workload conditions.

Automation misuse and disuse are problems associated with automation-interaction and, factors such as risk, workload, and self-confidence may affect misuse and disuse; under conditions of high-workload within which operators have little confidence in their capacity to
respond, misuse is a more likely outcome than disuse; misuse is associated to over-trust and disuse is associated to under-trust (Lee, 2008; Parasuraman & Riley, 1997). According to Kirlik (1993), burden associated to the use of a cognitive-aid can outweigh benefits from it; parameters of the multitask context and the cognitive-aid, such as engagement-times and disengagement-times, frequency and duration of secondary-task, and cost of delaying secondary tasks, determine whether task-offload aids result in improved performance of the operator; effort involved in the engagement and disengagement of the cognitive-aid can sometimes outweigh the benefits of the cognitive-aid (Kirlik, 1993). Dzindolet et al (1999) propose a model on automation usage based on data from four experiments; according to their model, automation use can be affected by perceived utility of the aid and perceived utility of the aid can be affected by trust-in-automation, and trust-in-self in being able to perform the task without automation assistance (Dzindolet, Pierce, Beck, & Dawe, 1999).

Upon interpreting results from this experiment, it appears that the withdrawal from the use of the Task-State Overview is possibly associated to the Workload-Preview not being perceived as useful as the perceived burden associated to its use outweighed perceived benefits resulting from its use especially under high Task-Load. Under high Task-Load some participants may have reached limits of mental resource capacity thereby resulting in the Task-State Overview going unused. Additionally, individuals may have regarded the Workload-Preview as redundant because current severity information provided sufficient information to switch between tasks, although through a smaller window of opportunity to do so. A fixed or ‘blind’ task-prioritization strategy of giving default priority to the primary task without scanning the Task-State Overview for verification was possibly followed by participants who reached maximum mental processing capacity under high Task-Load. This strategy that involved reduced verification was possibly developed under high task-load over time with exposure to several events in the scenario without the appearance of surge-events. A contributing factor to this was perceived complexity and higher
perceived mental-demand in processing Task-State Overview information. The development of such a strategy may have been possibly coupled with the development of the perception that the Task-State Overview was not useful. This probably led to withdrawal from visual scanning of the cognitive-aid in some individuals.

The prioritization of surge-events was affected by the use of the Workload-Preview. In the high Task-Load condition (i.e., Scenario D), the median Surge-Prioritization Error was 3 in the Workload-Preview Condition, and 2 in the Non Workload-Preview condition. In the Workload-Preview condition, more individuals possibly used a strategy that involved minimal scanning of Task-State Overview information as discussed above – a form of automation disuse. The effect of this is observed in a more spread-out distribution of Surge-Prioritization Errors in the Workload-Preview condition (refer to Figure 5-2). According to Wickens, in studies that concern responses to unexpected events, in the context of transportation safety, quantities such as statistical significance and differences (in mean values) between conditions are not as important as looking at the “absolute length and distribution” of observations (Wickens, 2001); even though Wickens stated this within the context of individual response-times, it is useful in the interpretation of Surge-Prioritization Errors (more than one responses/errors) as well, because extreme cases of Surge-Prioritization Error are a result of several individual surge-events missed repeatedly within the same scenario by one person. According to Wickens (2001), cases of human failure or error that compromise safety are a result of several factors such as high time-pressure, bad weather, low skill and unexpected events. In this study, instances of high Surge-Prioritization Error were a result of a combination of high Task-Load; low perceived utility of predictive information in the cognitive-aid; and individual differences in Task-Management strategy.
Implications to Cyber-Security

This dissertation explores a Human Factors issue within the context of cyber-security monitoring. The results obtained on various measures and lessons learned from this research could benefit future research exploring Human Factors issues in cyber-security especially within task-management. The complexity of organizational networks and increased sophistication of attacks calls for focused attention from the Human Factors community to address issues in task-prioritization in cyber-security monitoring and network surveillance. NETS-DART, the simulation framework utilized in this study or an extension of NETS-DART, may be used as a platform for exploring the effects of visualizations and attention-guidance technologies. NETS-DART is an example of a framework that goes beyond the typical Interrupting-Task/Ongoing-Task framework (DD Salvucci, Taatgen, & Borst, 2009) used to study behavioral issues in task-management and the effects of technology within that limited context. On complex tasks with dynamically changing conditions such as task-load and relative priorities of various sub-tasks where attention-management issues are a concern, the distinction between interrupting-tasks and ongoing-tasks is blurred (Wickens & McCarley, 2008), and exploring such problems with the Interrupting-Task/Ongoing-Task paradigm may not lead to generalizable results. The task-performance measure (i.e., Normal-Score), the concept of using multiple locations and multiple sub-tasks, Overall-Severities linked to the sub-tasks (to assign artificial priorities), the metaphor of a hypothetical organization with several departments, event-categories and priority-level definitions, and the process of constructing the task within a new context different from that of the predecessor to NETS-DART (i.e., emergency response) serves as contribution that may benefit future cyber-security research.
Contributions

Implications to Human Factors and Cyber-Security research discussed in previous subsections cover in detail some of the contributions that result from this dissertation. Additionally, the simulation (i.e. NETS-DART) and the measure of task-performance, although derived from previous work on NeoCities and idsNETS, are applied to a unique task context in this study. The demonstration of their use in this task context serves as a reference for their potential use in similar task contexts in the future. Event design, that encompasses the training material provided to identify Viruses, Worms, Vandalism Attack and Espionage Attack categories and priority-levels, would serve as a reference for future experimental research in Cyber-Security with undergraduate-student participants. Furthermore, Scenario-Design in this experiment that addresses task-load and surge-event manipulations in this study may guide potential experimental designs in the future that may involve similar experimental manipulations. Table 6-1 highlights contributions from this study.

Table 6-1. Set of contributions from the study

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Main Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusions</td>
<td>Implications to Workload-Previews and Attention-Guidance.</td>
</tr>
<tr>
<td>NETS-DART</td>
<td>• Platform for cyber-security research and testing.</td>
</tr>
<tr>
<td></td>
<td>• Platform for Task-Management or Dual-Tasking experimental research.</td>
</tr>
<tr>
<td>Scenario Design</td>
<td>• Implementation of Task-Load as a Factor.</td>
</tr>
<tr>
<td></td>
<td>• Implementation of Surge-Events or Rare/Unexpected Events especially within a secondary-task.</td>
</tr>
<tr>
<td></td>
<td>• First implementation of concurrent presentation of events.</td>
</tr>
<tr>
<td>Use of Subjective-Workload Measure</td>
<td>• Observed differences in Normal-Score between Moderate and High Task-Load scenarios provide implications on the extent to which scenarios varied in difficulty resulting from induced time-pressure. As the Subjective-Workload measure (TLX-Score) was significantly different between Moderate and High Task-Load scenarios, the values provide an estimate of scenario difficulty resulting from time-pressure. This measure could be used for further studies to calibrate scenario difficulty for experiments.</td>
</tr>
<tr>
<td>Cyber Event-Design</td>
<td>• Cyber-Security Decision-Making task that could be trained to most participants within 40 mins including quiz and training scenario.</td>
</tr>
</tbody>
</table>
| Normal-Score       | • Task-Performance was normally distributed.  
|                   | • This variable was effective as a Task-Performance measure.  
|                   | • As effects of learning were present with the Normal-Score, this measure could be used for training to criterion. An operator can be trained until the Normal-Score reaches a plateau.  
| Literature-Review | Task-Management, Task-Prioritization Literature, and Workload-Preview Literature that can be extended for other individual/team research in HCI. |

**Limitations**

From the results of the experiment (see page no. 89), it appears that surge-events were predictable by participants. A factor contributing to the ability to predict the appearance of secondary-task surge events may have been the timing of these surge-events. They were placed at the beginning of event batches in scenarios C and D. A random placement of surge-events in the secondary-task could have made them less predictable. It is also possible that the number of surge-events provided per participant may have been too many (i.e., 5 totally). The placement of unexpected events in a simulation is suggested to be limited to one for some studies (Wickens, 2001). Based on the discussion on Task-Prioritization Error (see page no. 89) and the effects of learning (see page no. 96), it appears that following the prescribed task-prioritization strategy was relatively easy for participants. A relatively more complex task-prioritization strategy may have resulted in normal distributions of task-prioritization error. The use of the Workload-Preview within the context of more complex task-prioritization strategy may have resulted in more findings applicable to dual-task event monitoring. Another limitation from the study was that the event-volume component in the Task-State Overview may have not been very useful within the context of the scenarios developed in this experiment. Although participants may have used the event-volume component to track incoming events or to track events recently added by comparing event volume history, it is not clear if this was critical to task-management. This limitation regarding the usefulness of event-volume is especially worth noting given that the
hypotheses were not supported. The Workload-Preview provides a 12 second look-ahead for a change in sub-task relative priorities. The study does not address implications of memory and rehearsal pertaining to the use of the Workload-Preview and increased mental-workload owing to working-memory usage. Future studies could address implications of Workload-Preview on memory limitations. Manipulation check items should have been introduced to verify whether participants could predict surge-events. Another useful manipulation check that could have been introduced is a question that would ask the participant to rate the extent to which they scanned the task-state overview at different stages such as - after completing events, before beginning to read event descriptions, etc.

**Future Work**

The Workload-Preview considered in this study was perfectly reliable. While it is necessary to investigate the effects of perfectly reliable systems on performance, many technological aids do not possess perfect reliability. Future work may involve investigating the effect of Workload-Previews with false-positives. Introducing false-positives within the technological aid being tested can involve a more complex scenario design and was therefore not considered within this study. Moreover, testing for effects of false-positives require a perfect reliability baseline in addition to several levels of false-alert rates to compare.

Real-world event-monitoring often involves multiple tasks with more complex prioritization strategies for attending to sub-tasks. The use of more than two homogeneous and heterogeneous sub-tasks (Raby & Wickens, 1994; Segal & Wickens, 1990) within the cybersecurity event-monitoring context involving complex prioritization strategies, would provide interesting results on the effects of Workload-Previews on task-performance.
References


Appendix A

Definition of Event-Categories and Priority-Levels

The following event definitions were provided to participants during training:

Identification of a Virus

- **Virus**: A virus is malicious code that operates and spreads by altering the code of the host file.
- **Main Risks**: Damage to important or critical software (!) in use to the point it becomes unusable.
- **Signs in your report that indicate a Virus:**
  1. Unexpected messages from screen or odd sounds from speaker.
  2. One or more infected programs have been reported.
  3. Applications may show erratic behavior (e.g. MS Word would open unwanted files).
  4. File-Integrity Loss: time/date stamps on files may be changed; some program or file sizes increase substantially; some files may have double extensions (such as filename.pdf.exe).

Identification of a Worm

- **Worms**: Worms unlike viruses do not need to attach themselves to other programs to spread or modify other programs. Worms do not infect other programs, but only use their functionality.
• **Main Risks:** Left unchecked, worms can spread throughout the network resulting in degraded performance and possible loss of service.

• **Signs in your report that indicate a Worm:**

  (1) Consumes an excess of system memory and/or network bandwidth.

  (2) There are multiple copies of an unusual file in the system and/or hard drives may have filled up rapidly.

  (3) Users lose access to directories OR to drives where access was previously available.

  (4) Users cannot access security software utilities or request updates.

**Common Properties of Viruses and Worms:**

(1) spread with email-attachments, downloads or portable memory such as flash-drives, external hard-drives and optical discs.

(2) cause system slow down, system crashes/reboots

(3) take advantage of computers that do not have the latest security updates installed.

**Identification of a Vandalism Attack**

• **A Vandalism attack** is activity conducted by an individual, on workstations or the organizational network, with an intent to destroy cyber infrastructure, delete files/software, or corrupt web sites to spoil company reputation. The motivation is to seek revenge or make a statement.

• **Goals of a Vandalism Hacker:**

  To delete files or programs important to organizational functioning.

  To deface company website pages viewable to public.

  To send out emails to people on address books, with content detrimental to the company.
• **Signs in your report that indicate a Vandalism Attack:**
  
  (1) Hacker activity is random in terms of *when* during the day they act.
  
  (2) Files or documents may have been deleted.
  
  (3) Evidence of revenge-motive or a “disturbing message” may be present somewhere.
  
  (4) Changed system settings, network settings or any evidence of casual tampering.

**Identification of an Espionage Attack**

• **An Espionage attack** is activity conducted by an individual, with an intent of spying or exfiltrating intellectual property or classified information.

• **Goal of Espionage Hacking** – To get access to secret documents and exfiltrate it.

• **Signs in your report that indicate an Espionage Attack:**
  
  (1) Microphones, key-logger, cameras may be turned on for surveillance without the user knowing.
  
  (2) Communication from a host computer from a known hostile country (e.g. Iran) or a rival company.
  
  (3) Evidence of social engineering – unknown intruders got past gate security or trusted individuals turned out to be spies!
  
  (4) Hacker was searching around folders that may possibly contain classified documents; Search keywords of interest may be left behind.

**Common Properties of Vandalism Attacks and Espionage Attacks:**

(1) Unusual network activity in Intrusion Detection Systems (IDS’s) or firewalls. (Note that it is not necessary for you to know what an IDS does and alerts related to IDS’s will be explicitly identified)

(2) Multiple (failed) log-in attempts into company user accounts.
Appendix-Figure 1. Instructions to participants for assigning priority-levels to Viruses and Worms

<table>
<thead>
<tr>
<th>CATEGORIES</th>
<th>PRIORITY 1</th>
<th>PRIORITY 2</th>
<th>PRIORITY 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>•Virus</td>
<td>Benign</td>
<td>Malignant</td>
<td>Malignant and Infection has Spread to Critical Software</td>
</tr>
<tr>
<td></td>
<td>• Signs of a virus are present.</td>
<td>• Some programs have stopped working.</td>
<td>• Software used for work purposes is not functional.</td>
</tr>
<tr>
<td></td>
<td>• System may be slowed.</td>
<td>• Software used for work purposes is functional.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• All programs work.</td>
<td>• Currently contained within one workstation.</td>
<td></td>
</tr>
<tr>
<td>•Worm</td>
<td>Benign</td>
<td>Malignant</td>
<td>Malignant and Spreading to Network</td>
</tr>
<tr>
<td></td>
<td>• Signs of a worm are present.</td>
<td>• Software used for work purposes is not usable.</td>
<td>• Spreading to neighboring systems within the same department or other departments.</td>
</tr>
<tr>
<td></td>
<td>• System may be slowed.</td>
<td>• Currently contained within one workstation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• All programs work.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GUIDELINES FOR PRIORITY LEVEL ASSIGNMENT – Virus and Worm

Appendix-Figure 2. Instructions to participants for assigning priority-levels to Vandalism and Espionage events

<table>
<thead>
<tr>
<th>CATEGORIES</th>
<th>PRIORITY 1</th>
<th>PRIORITY 2</th>
<th>PRIORITY 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>•Vandalism Attack</td>
<td>Attempted Only</td>
<td>Some Vandalism Done</td>
<td>Goal Attained</td>
</tr>
<tr>
<td></td>
<td>• Hacking Attempted.</td>
<td>• Some vandalism identified.</td>
<td>• Critical software, documents, deleted.</td>
</tr>
<tr>
<td></td>
<td>• No vandalism yet.</td>
<td>• Critical software/documents remain safe.</td>
<td>• Externally facing website defaced/spoiled.</td>
</tr>
<tr>
<td></td>
<td>• Indications of motivation to vandalize are present.</td>
<td>• Critical software or internet may have slowed down but is usable.</td>
<td>• Unauthorized emails have been sent that have negative impact on the company.</td>
</tr>
<tr>
<td>•Espionage Attack</td>
<td>Attempted Only</td>
<td>Evidence of Some Documents Accessed</td>
<td>Goal Attained</td>
</tr>
<tr>
<td></td>
<td>• Hacking or Social Engineering Attempted.</td>
<td>• Information not publicly available has been accessed.</td>
<td>• Classified documents or intellectual-property accessed.</td>
</tr>
<tr>
<td></td>
<td>• Any information accessed is publicly available.</td>
<td>• Classified documents or intellectual-property have not been accessed.</td>
<td>• Classified documents or intellectual property are copied and transferred to a destination.</td>
</tr>
</tbody>
</table>
Appendix B: Training Quiz Questions

Based on the material provided to you in the training materials, please answer the following questions. You may take approximately 10 mins to answer them. You are allowed to refer to the hard-copy slides provided to you. Each question has one answer among the options listed. Please select them accordingly. If you have any questions, please let the experimenter know.

Q 1 Changed network settings in a server and disturbing email messages sent out to a number of clients, would be sufficient to file a report for -
○ An Espionage Attack
○ A Vandalism Attack
○ A Worm
○ A Virus

Q 2 Which of the following conditions is sufficient to report a system with a Worm, with level-3 priority?
○ Important software is use (e.g. MS Excel) does not work
○ The Worm has started to spread to other machines
○ System is crashing frequently
○ Too much system memory is being used up

Q 3 File-integrity loss, or specifically, significant changes in file sizes not resulting from user action, is a sign of a Worm.
○ True
○ False

Q 4 A trespasser somehow walks past gate security, illegally obtains access to a network server, copies intellectual property stored in it, and walks out. This should be reported as:
○ Virus, Priority 3
○ Worm, Priority 3
○ Espionage, Priority 3
○ Espionage, Priority 2

Q 5 A user downloads a file attachment from an email. The system has slowed down considerably since then. The system does not have the latest security updates. This indicates –
○ The presence of a Worm, but no Virus
○ The presence of a Virus, but no Worm
○ The presence of a Worm or a Virus
○ It is definitely not a Virus or Worm

Q 6 A hacker obtains unauthorized access to a system that hosts a database that is critical for conducting transactions. The hacker attempted to disable the database without success. However,
other folders and files containing unimportant old records were deleted. This should be reported as:
- Vandalism, Priority 2
- Vandalism, Priority 3
- Espionage, Priority 2
- Espionage, Priority 3

Q 7 A Hacker located in North Korea, has been repeatedly attempting to obtain unauthorized access to the website server. The hacker was not successful in accessing anything not available to the general public. This should be categorized as:
- Espionage, Priority 2
- Vandalism, Priority 1
- Espionage, Priority 1
- Vandalism, Priority 2

Q 8 A system contains a Worm. The system is mainly used for developing animations for the marketing department. The worm has prevented the use of the animation software. The priority level for reporting this is:
- Priority 1
- Priority 2
- Priority 3
- The information is insufficient to determine the priority level.

Q 9 A system contains a Virus. The most critical software on this system is Eclipse used in software-development. The virus has prevented the use of Eclipse. The priority level for reporting this is:
- Priority 1
- Priority 2
- Priority 3
- The information is insufficient to determine the priority level.

Q 10 Sam, an employee gets a phone-call from a person who falsely claims to be working in the IT department next door. Sam complies with a request to provide his username and password. This should be reported as an Espionage attack because:
- Signs of social-engineering are present
- Signs of casual tampering are present
- Search keywords of interest in sensitive documents are present
- None of these
Q 11 Suppose the Core Departments is selected in the Departments Monitor - Upon taking a look at the Task-State Overview, under which of the following conditions would you switch from the Core Departments to External Departments? (Note: you will NOT find this answer in the hard-copy slides and you may refer to the training slides on your screen)

STOP. PLEASE WAIT FOR THE EXPERIMENTER TO SUBMIT YOUR RESPONSES.
## Appendix C: Subjective Workload Questionnaire

Based on the FIRST PERFORMANCE SCENARIO (not training) you completed on the NETS-DART simulation, please rate the following:

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mental Demand:</strong> How mentally demanding was the task?</td>
<td>![Rating Scale]</td>
</tr>
<tr>
<td><strong>Physical Demand:</strong> How physically demanding was the task?</td>
<td>![Rating Scale]</td>
</tr>
<tr>
<td><strong>Temporal Demand:</strong> How hurried or rushed was the pace of the task?</td>
<td>![Rating Scale]</td>
</tr>
<tr>
<td><strong>Performance:</strong> How successful were you in accomplishing what you were asked to do?</td>
<td>![Rating Scale]</td>
</tr>
<tr>
<td><strong>Effort:</strong> How hard did you work to accomplish your level of performance?</td>
<td>![Rating Scale]</td>
</tr>
<tr>
<td><strong>Frustration:</strong> How irritated, stressed, annoyed or frustrated were you?</td>
<td>![Rating Scale]</td>
</tr>
</tbody>
</table>
Appendix D: End-Task Survey

With reference to the 4 scenarios you completed in the NETS-DART simulation please respond to the following items:

Q1 Instability of Situation  How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straightforward (low)?
   LOW______________ HIGH.

Q2 Complexity of Situation  How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (Low)?
   LOW______________ HIGH.

Q3 Variability of Situation  How many variables are changing in the situation? Are there a large number of factors varying (high) or are there very few variables changing (low)?
   LOW______________ HIGH.

Q4 Arousal   How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?
   LOW______________ HIGH.

Q5 Concentration of Attention  How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?
   LOW______________ HIGH.

Q6 Division of Attention  How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focused on only one (low)?
   LOW______________ HIGH.

Q7 Spare Mental Capacity  How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (high) or do you have nothing to spare at all (low)?
   LOW______________ HIGH.

Q8 Information Quantity  How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?
   LOW______________ HIGH.

Q9 Information Quality  How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?
   LOW______________ HIGH.

Q10 Familiarity with the Situation  How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?
   LOW______________ HIGH.
POST ACTIVITY SURVEY

Q11 Please select the statement which best applies to you:
- The majority of games I play are computer games.
- The majority of the games I play are console games (i.e. PS3, Xbox, Wii, etc.)
- I play an equal amount of computer and console games
- I don’t play either computer or console games

Q12 How many HOURS A WEEK ON AVERAGE do you spend playing:

<table>
<thead>
<tr>
<th>Game Type</th>
<th>No. of Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time strategy games (Dawn of War, starcraft, command &amp; conquer, sim city, Civilization, etc.)</td>
<td></td>
</tr>
<tr>
<td>Role Playing Games (Mass Effect, Assassins Creed)</td>
<td></td>
</tr>
<tr>
<td>Multiplayer Online Role player Games (World of Warcraft, Age of Conan, Star Wars Galaxies, etc.)</td>
<td></td>
</tr>
<tr>
<td>Team-based First Person Shooter Games (Call of Duty, Team Fortress, Halo, etc)</td>
<td></td>
</tr>
<tr>
<td>Other Games</td>
<td></td>
</tr>
</tbody>
</table>

Q13 Have you ever played or taken part in an experiment using the NeoCITIES Emergency Response simulation
- Yes
- No

Q14 If the above is a 'Yes', please mention when:

Q15 Please check the boxes that relate to any previous experience working in IT (i.e., IT Support, Network Operations, Network Design, etc.)
- I have held a full-time job in IT
- I have held a part-time job in IT
- I have taken courses in IT/Computer Science in the last 4 years
- None of the above

Q16 Please select the option that relates most to the amount of experience you may have working in IT (i.e., IT Support, Network Operations, Network Design, etc.)
- None
- 1 – 3 years
- 3 – 5 years
- 5 – 10 years
- 10+ years
Q17 If the previous answer was not ‘None’, please describe your experience working in Information Technology.

Q18 Please rate your previous experience working with Intrusion Detection Data or other Cyber-Security analysis of log-data (1 – 7; 0 = No Experience, 7 = Highly Experienced).
   No Experience ____________________________________________________ Highly Experienced.

Q19 If you have worked with Intrusion Detection Software previously, please list software that you may have used:

Q20 Have you taken a computer networks class in the last 4 years? (Yes/No). If Yes, please specify the level or a specific course-number.

Q21 Have you taken a cyber-security class in the last 4 years? (Yes/No). If Yes, please specify the level or a specific course-number.

Q22 Based on your experiences in the simulation, how easy were each of the interface components for you to understand? (1-Very Difficult to understand, 5 – Very Easy to understand)

<table>
<thead>
<tr>
<th></th>
<th>1 – Very Difficult</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 - Very Easy</th>
<th>I do not know what this is</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department Monitor</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Report Description</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Feedback Panel</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Categorization Panel</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Task State Overview</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Q23 Based on your experience how useful do you think each of the following components are to completing your task (1- Completely Useless, 5- Very Useful)

<table>
<thead>
<tr>
<th>Component</th>
<th>1 – Completely Useless</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 – Very Useful</th>
<th>I do not know what this is</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department Monitor</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Report Description</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Feedback Panel</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Categorization Panel</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Task State Overview</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Q24 During the experiment, did you have any problems or confusions that impeded your ability to complete the task? Please describe.

Q25 If you were designing training materials to use this system, what would you include?

Q26 Based on your experience please answer the following questions based on how much you agree with the statement (1- Completely Disagree, 5 Completely Agree).

<table>
<thead>
<tr>
<th>Question</th>
<th>1 - Completely Disagree</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 - Completely Agree</th>
<th>I do not understand what this means</th>
</tr>
</thead>
<tbody>
<tr>
<td>I was able to successfully complete the task.</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I had an adequate amount of time to complete the task.</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>During the simulation, I was aware of the situation at all departments.</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I was able to understand and utilize the feedback</td>
<td>○</td>
<td>○</td>
<td></td>
<td></td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Provided to me in the Feedback Panel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>The predictive information provided by the Task-State Overview was helpful in deciding when to switch between the primary and secondary tasks.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Once I found an event, I was able to figure out how to file a report about it.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I was able to report events effectively.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I was able to keep track of my actions in the system.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>In the Department Monitor, I found the color of the circle useful in selecting my next location.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I used the Task-State Overview during my task.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The Task-State Overview was easy to</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
The Task-State Overview did not interfere with my interaction with the task.

The predictive information provided by the Task-State Overview was confusing.

I could have performed well without the predictive information in the Task State Overview.

| Q27 | You may have observed that new reports came in 'batches'. Did you notice new incoming batches of reports, while working on a report in the EXTERNAL Department, OR, after you just finished working on a report in the EXTERNAL Department?
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes, but not all the time</td>
<td>I do not remember, or, I am not sure</td>
</tr>
</tbody>
</table>

| Q28 | If the answer to the previous question is NOT 'NO', did this help you in giving higher priority to reports in the External Department in situations where the Overall-Severity of the External Departments was higher than the Overall-Severity of the Core Departments?
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes, but not all the time</td>
<td>I do not remember, or, I am not sure</td>
</tr>
</tbody>
</table>

| Q29 | If there is anything you wish you knew before the simulation, what would it be? |
Q30 Please rate the extent to which you agree or disagree with the following statements about yourself, on a 5 point scale (1= Strongly Disagree, 5 = Strongly Agree):

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I prefer to work on several projects in a day, rather than completing one project and then switching to another.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would like to work in a job where I was constantly shifting from one task to another, like a receptionist or an air traffic controller.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I lose interest in what I am doing if I have to focus on the same task for long periods of time, without thinking about or doing something else.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When doing a number of assignments, I like to switch back and forth between them rather than do one at a time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I like to finish one task completely before focusing on anything else.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It makes me uncomfortable when I am not able to finish one task completely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>before focusing on another task.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am much more engaged in what I am doing if I am able to switch between several different tasks.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I do not like having to shift my attention between multiple tasks.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I would rather switch back and forth between several projects than concentrate my efforts on just one.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I would prefer to work in an environment where I can finish one task before starting the next.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I don’t like when I have to stop in the middle of a task to work on something else.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>When I have a task to complete, I like to break it up by switching to other tasks intermittently.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I have a “one-track” mind.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I prefer not to be interrupted when working on a task.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1. Disagree strongly  
2. Disagree a little  
3. Neither agree nor disagree  
4. Agree a little  
5. Agree strongly

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1 - Disagree Strongly</th>
<th>2 - Disagree a little</th>
<th>3 - Neither agree nor disagree</th>
<th>4 - Agree a little</th>
<th>5 - Agree strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is talkative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tends to find fault with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does a thorough job</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is depressed, blue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is original, comes up with new ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is helpful and unselfish with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can be somewhat careless</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is relaxed, handles stress well</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is curious about many different things</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is full of energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starts quarrels with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is a reliable worker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can be tense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is ingenious, a deep thinker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generates a lot of enthusiasm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has a forgiving nature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Tends to be disorganized</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Worries a lot</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Has an active imagination</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Tends to be quiet</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is generally trusting</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Tends to be lazy</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is emotionally stable, not easily upset</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is inventive</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Has an assertive personality</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Can be cold and aloof</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Perseveres until the task is finished</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Can be moody</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Values artistic, aesthetic experiences</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is sometimes shy, inhibited</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is considerate and kind to almost everyone</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Does things efficiently</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Remains calm in tense situations</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Prefers work that is routine</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is outgoing, sociable</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Is sometimes rude to others</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Makes plans and follows</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>through with them</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Gets nervous easily</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Likes to reflect, play with ideas</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Has few artistic interests</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Likes to cooperate with others</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Is easily distracted</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Is sophisticated in art, music, or literature</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Is politically liberal</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Please fill the following demographic information

**Q32** Please enter your Academic Major:

**Q33** Gender
- Male
- Female

**Q34** Age

**Q35** Your first language

**Q36** Year in Academic Program
- Freshman
- Sophomore
- Junior
- Senior
- Graduate Student
- Other
Q37 Race (optional)
- Asian
- Black/African American
- Hispanic
- Middle Eastern
- Native
- South-Asian
- White/Caucasian
- Other

Q38 Course from which you were recruited for this experiment

Q39 Please mention any color blindness problems you may have (optional):

DEBRIEFING
The experiment was primarily designed to investigate the effect that task-load and a workload-preview (presented as the Task-State Overview) have on the ability to prioritize between sub-tasks. The situations or events presented within this task, were not necessarily designed to emulate reality but to create scenarios and events in this scaled-world experiment. The specific definitions of Worms, Viruses, Intrusion for Espionage, or Intrusion for Vandalism provided to you are not accurate, but, were simplified for this experiment. These information security threats are broad classifications and their exact nature may vary from instance to instance of the threat.
VITA

Dev Minotra

Email: dxm401@psu.edu,
Phone: 814-206-4354

Education

The Pennsylvania State University, University Park.
Ph.D. Information Sciences and Technology, Expected 2012
University of Windsor, Canada.
University of Pune, India.

Publications


