MODELING HUMAN ATTENTION AND PERFORMANCE IN AUTOMATED ENVIRONMENTS WITH LOW TASK LOADING

by

Fei Gao

B.S. Industrial Engineering, Tsinghua University, 2008
M.S. Industrial Engineering, Tsinghua University, 2010

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

Author: ____

Fei Gao, Institute for Data, Systems, and Society
January 29, 2016

Certified by: ____

John Carroll
Gordon Kaufman Professor of Management
Professor of Organization Studies and Engineering Systems
Thesis Supervisor

Certified by: ____

M. L. Cummings
Visiting Professor
Aeronautics and Astronautics

Certified by: ____

Nelson Repenning
School of Management Distinguished Professor of System Dynamics and Organization Studies

Accepted by: ____

John N. Tsitsiklis
Clarence J. Lebel Professor of Electrical Engineering
Graduate Officer, Institute for Data, Systems, and Society

MODELING HUMAN ATTENTION AND PERFORMANCE IN AUTOMATED ENVIRONMENTS WITH LOW TASK LOADING

by
Fei Gao

Submitted to the Institute for Data, Systems, and Society
on January 29, 2016 in partial fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Engineering Systems Division

Abstract

Automation has the benefit of reducing human operators' workload. By leveraging the power of computers and information technology, the work of human operators is becoming easier. However, when the workload is too low but the human is required to be present either by regulation or due to limitations of automation, human performance can be negatively affected. Negative consequences such as distraction, mind wandering, and inattention have been reported across many high risk settings including unmanned aerial vehicle operation, process control plant supervision, train engineers, and anesthesiologists. Because of the move towards more automated systems in the future, a better understanding is needed to enable intervention and mitigation of possible negative impacts.

The objectives of this research are to systematically investigate the attention and performance of human operators when they interact with automated systems under low task load, build a dynamic model and use it to facilitate system design. A systems-based framework, called the Boredom Influence Diagram, was proposed to better understand the relationships between the various influences and outcomes of low task loading. A System Dynamics model, named the Performance and Attention with Low-task-loading (PAL) Model, was built based on this framework. The PAL model captures the dynamic changes of task load, attention, and performance over time in long duration low task loading automated environments.

In order to evaluate the replication and prediction capability of the model, three dynamic hypotheses were proposed and tested using data from three experiments. The first hypothesis stated that attention decreases under low task load. This was supported by comparing model outputs with data from an experiment of target searching using unmanned vehicles. Building on Hypothesis 1, the second and third hypotheses examined the impact of decreased attention on performance in responding to an emergency event. Hypothesis 2 was examined by comparing model outputs with data from an experiment of accident response in nuclear power plant monitoring. Results showed that performance is worse with lower attention levels. Hypothesis 3 was tested by comparing model outputs with data from an experiment of defensive target tracking. The results showed that the impact of decreased attention on performance was larger when the task was difficult. The process of testing these three hypotheses shows that the PAL model is a generalized theory that could explain behaviors under low task load in different supervisory control settings. Finally, benefits, limitations, generalizability and applications of the PAL model were evaluated. Further research is needed to improve and extend the PAL model, investigate individual differences to facilitate personnel selection, and develop system and task designs to mitigate negative consequences.

Thesis supervisor: John Carroll
Title: Gordon Kaufman Professor of Management, Professor of Organization Studies and Engineering Systems

Acknowledgements

In investigating boredom, I found a very useful theory on the optimal experience of flow, which is a feeling of energized focus, full involvement, and enjoyment in the process of an activity. Achieving flow needs a clear set of goals and progress to add direction and structure to the task, as well as clear and immediate feedback, and a good balance between the perceived challenges of the task at hand and a person's own perceived skills. In one period of my life, I lacked all three. I was stuck endlessly waiting, had no direction, and could do almost nothing about the situation. I experienced a mental state different from flow and boredom: anxiety and depression. I owe thanks to numerous people for the successful completion of this thesis. Without them, I may not have even had a chance to write this thesis.

To my advisor John Carroll, for taking me as your student and advising me on the two research topics I worked on. You helped me when I felt hopeless and almost gave up. Without your help, I would not have been able to return to MIT to continue my study.

To my former advisor Missy Cummings, for always supporting me over this long journey. Being able to join your lab was a surprise for me. You provided me such a great opportunity to learn and grow. Thank you for all the advice and encouragement you gave and your detailed editing for my writings. What's even more substantial is that you suffered through my most difficult years together with me. You came up with new suggestions to help me when previous ones failed. This is more than what most advisors would do. For all of these, I will always be grateful.

To Nelson Repenning, thanking you for advising me on the techniques and art of System Dynamics modeling, which is critical for improving the quality of my model.

To my fellow HALiens: Christin S. Hart, Kris Thornburg, Mark Boyer and Jamie Macbeth, thank you for sharing your experiment data to be used in this thesis. Erin Solovey, thank you for your valuable advice on research and life. Luca Bertuccelli, thank you for being my mentor in the lab and in UTRC. Yves, Jason, Alex, Kim, Armen, Paul, Kathleen, and many others, thank you for the constant support, as well as all of the good times. I missed the days in HAL after I became a "lonely wolf".

To Elizabeth Milnes, for your advice and support in each step of my academic life in ESD. Thank you for reassuring me when I called you from China, crying in the middle of the night.

To Danielle Guichard-Ashbrook and other staff of the International Students Office, for all the work you did to handle my special case. Danielle, hope you will rest in peace.

To Sally Chapman, for all the support and help that you've provided through my years at HAL. Thank you for getting my mat signed as the last student in HAL.

To Murilo Bonilha, Jinlei Ding, Mei Chen and my other colleges at UTRC – China, for the valuable work experience, great times, and kindly help. Mei, thank you for all the coffees and your wise advice. I will remember to think hard.

To Ying Wang, thank you for sharing your home with me. Gaofeng Sun, thank you for being my personal trainer. Exercise, the one thing that I could see progress during the endless waiting, prevented me from deep depression. To many other friends, thank you for providing help and companions in my most difficult days.

To my parents, for your love and support over these many years. You know nothing about my research, but you know much about when to motivate, encourage, or comfort me. You did not have the opportunity when you were young, so you did everything you could to make sure I could get a good education. I love you, and this is all dedicated to you.

To Jingyu Tan, thank you for encouraging and supporting me to follow my dream. Although we are far in distance, we are close in heart. Your parents are warm-hearted people who offered a lot of help. I cannot wait to finally moving together and starting a life with you.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ADHD	Attention Deficit Hyperactivity Disorder
ANOVA	Analysis of Variance
BID	Boredom Influence Diagram
BPS	Boredom Proneness Scale
ECR	Executive Control Resource
ECG	Electrocardiogram
EEG	Electroencephalogram
fNIRS	Functional Near Infrared Spectroscopy
HOMER	Human Operator Monitoring Emergent Reactor
MSBS	Multidimensional State Boredom Scale
MSE	Mean Square Error
NASA TLX	National Aeronautics and Space Administration Task Load Index
NEO PI	Neuroticism-Extraversion-Openness Personality Inventory
NEO-FFI-3	NEO Five-Factor Inventory-3
OPS-USERS	Onboard Planning System for UxVs Supporting Expeditionary Reconnaissance and Surveillance
PAL	Performance and Attention with Low-task-loading
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
SCT	Schedule Comparison Tool
SD	System Dynamics
S.D.	Standard Deviation
TUT	Task Unrelated Thought
U^C	Covariation component of MSE
U^M	Bias component of MSE
U^S	Variation component of MSE
UAV	Unmanned Aerial Vehicle

1 Introduction

As early as the 19th century with the pending industrial revolution, Nietzsche (1878) warned that a machine culture would cause boredom for workers, resulting in human "play" at work. More than 130 years later, the news is replete with examples of just such a phenomenon. In 2009, during the enroute portion of a flight, two Northwest pilots reportedly were distracted by their laptops causing them to overfly Minneapolis by 90 minutes ("Northwest airlines flight 188" 2009). In 2011, an air traffic controller and a supervisor were suspended after it was discovered that the controller was watching movies in the early morning hours under reported light traffic conditions ("Movie-watching air traffic controller suspended" 2011). While ultimately distraction was the direct cause of operator misbehavior in these cases, the under-stimulating low task load, i.e., boring, environment was a clear contributing factor.

Boredom and associated serious negative consequences have been reported across many other high risk settings including unmanned aerial vehicle operation (Thompson et al. 2006), process control plant supervision (Sheridan et al. 1983), train engineers (Haga 1984), train drivers, and professional truck drivers (Dunn and Williamson 2011; Oron-Gilad et al. 2008), as well as anesthesiologists (Weinger 1999). In boring environments where task load is low, typical in highly automated supervisory control environments, operators often find other tasks to help them sustain some level of attention and in many cases, simply to help them stay awake. With a global push to introduce more automation and autonomy into numerous safety-critical work environments (e.g., driverless cars, positive train control in rail, and completely automated mines), boredom will likely be a growing problem.

While boredom in safety-critical settings is of obvious concern, it is also pervasive across more benign office work environments, often with such negative consequences as absenteeism and poor retention (Fisher 1993). The Internet is replete with articles, blogs, and forums providing advice on how to survive and cope with a boring job. New social media sites such as glassdoor.com and indeed.com have emerged that allow employees the ability to anonymously rate their work environment, and comments such as "quite boring work environment with a lot of overtime" and "satisfactory but boring" are commonplace.

In 2006 in the UK, 2,113 college graduates aged 21 to 45 were surveyed about workplace boredom, with 61% reporting boredom due to the lack of a challenging job. Those in administrative and manufacturing jobs reported the highest boredom, while healthcare workers and teachers reporting the least boredom ("Teaching 'the least boring job'" 2006). Boredom in the workplace has been identified as an important issue in organizational research (Fisher 1993; Loukidou et al. 2009).

Research has shown that boredom is often associated with significant health problems. Boredom has been linked to premature death due to cardiovascular disease (Britton and Shipley 2010), and has been given as a primary reason for recreational drug use (McIntosh et al. 2005). Boredom proneness has been linked to increasing risk of anxiety and depression (Sommers and Vodanovich 2000; Vodanovich et al. 1991), as well as substance abuse (Farmer and Sundberg 1986; LePera 2011) and eating disorders (Abramson and Stinson 1977).

Given the ubiquity of boredom across a wide spectrum of work environments, exacerbated by increasing automated systems and advanced technologies (Nietzsche's forewarned "machine culture"), which remove humans from direct, physical system interactions and possibly increasing tedium in the workplace, there is a need to not only better understand the multiple facets of boredom in work environments, but to develop targeted mitigation strategies. To better understand the relationships between the various influences and outcomes of boredom, models of various elements of boredom and their interrelationships are needed.

1.1 Motivation

While the issues of boredom in the workplace in general, and more specifically in highly automated environments, are known to researchers and practitioners, they have generally not been as well researched as other areas such as vigilance and the effect of high workload on performance. The goal of many automation system designers is to achieve full automation and take humans out of the loop. When this is not possible, humans are placed in the system as a supervisor and a backup to solve problems beyond the capability of automation. This view not only ignores the value of humans, but also poses a high requirement on attention and perception over long duration time periods, leading to degraded performance and safety concerns. The short-term effect of distraction, loss of situation awareness, and the long-term effect of skill degradation all contribute to such negative consequences. The first step to solve such issues is to organize the various facets of boredom and investigate human behavior in long duration, low task load human automation interaction tasks.

Secondly, because of the move towards more automated systems in the future, a better understanding of environments that lead to boredom and associated negative behaviors is needed to enable intervention and mitigation of these possible negative impacts. A recent study claimed that the continuous vigilance approach of many current real-world monitoring tasks is doomed to fail (Casner and Schooler 2015). To create an optimal experience, also called flow experience, one must be involved in an activity with 1) a clear set of goals and progress, 2) clear and immediate feedback, and 3) a good balance between the perceived challenges of the task at hand and their own perceived skills (Csikszentmihalyi et al. 2014). However, many current human-automation systems lack all three of these conditions.

Human operators are often required to monitor for alarms and emergency event, but are unclear about what could happen and when. For most of the time, automation is capable of handling the tasks and there is not much left to do by the human, resulting in underload and underutilization in terms of both work time and human intelligence. This calls for changes in task design, system management strategies, and even a paradigm shift when we consider the human-automation interrelationship.

The introduction of more automation in complex systems means that boredom once caused by monotonous and repetitive tasks is now shifting to boredom caused by low task loading in the monitoring of such systems. And while there is significant previous work in the relationship of boredom to monotonous and repetitive environments, there is a paucity of research on work environments that address human behavior and possible mitigations in environments with almost nothing to do, both with and without highly automated systems (Fisher 1993).

To these ends, a systems-based framework was proposed that describes various elements of boredom and their interrelationships. This work will present a system dynamics model, which captures the dynamic changes of task requirement, attention, and performance overtime. This model can also be used to assess the impact of automation and system design alternatives on the whole system with human in the loop.

Simulation models are valuable in capturing the process and dynamics of human-automation interaction. As shown in Figure 1-1, a simulation model could build a connection between a conceptual model of the real world and solutions. With a valid simulation model, we can test and compare proposed changes to the current system, or new designs of the system at a lower cost than testing directly in the real world. In addition, with simulation models, we can learn about the causal relations among system components faster than learning from real world experience. My research aims to investigate the attention and performance of human operators when they interact with automated systems under low task load.

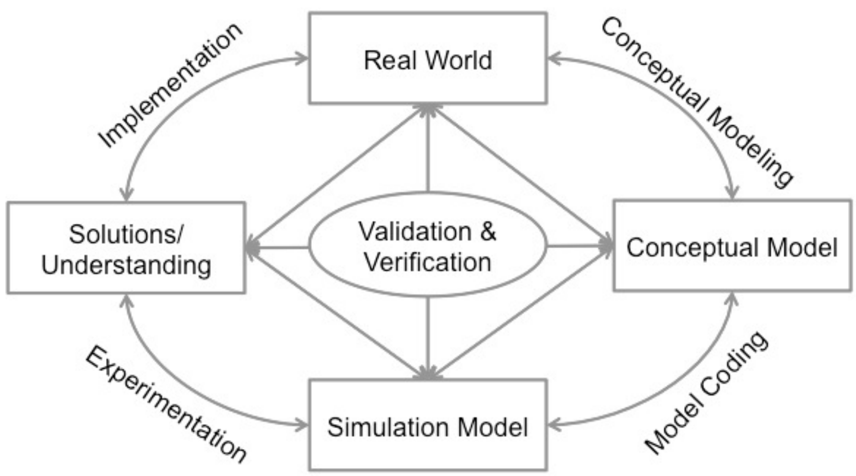


Figure 1-1: Research Approach (Sterman 2000)

1.2 Research Questions

The objectives of this research are to systematically investigate the attention and performance of human operators when they interact with automated systems under low task load, build a dynamic model and use it to facilitate system design. These objectives lead to the following three research questions:

- What are the major factors and influences that affect boredom of human operators when they interact with automated systems? Which of these factors and influences should be captured in a model?
- Can such a model be used to predict the impact of automation and boredom coping strategies on human and system performance?
- What level of accuracy can be expected of this model? What are the boundary conditions of the model?

1.3 Thesis Organization

In order to address these research questions, this thesis has been organized into the following chapters:

- Chapter 1, *Introduction*, describes the motivation, objectives, and research questions of this thesis.
- Chapter 2, *A Systems View of Boredom: The Boredom Influence Diagram*, reviews historical and recent efforts in boredom research and related fields. It introduces a systems-based framework, called the Boredom Influence Diagram, which describes various elements of boredom and their interrelationships. Mathematical and theoretical models related to boredom are also reviewed.
- Chapter 3, *Model Development*, describes the modeling process that created the Performance and Attention with Low-task-loading (PAL) model, a System Dynamics (SD) model of human-automation interaction in long duration, low task load scenarios. The chapter describes the model in detail, proposes three dynamic hypotheses regarding human attention and performance, and presents the results of model structure tests.
- Chapter 4, *Hypothesis 1: Hours of Boredom*, describes a human subject experiment that evaluates the ability of the PAL model to replicate the decrease of human attention and performance in a low task-loading scenario. The effectiveness of attention alerting is predicted. The quantitative predictions made by the PAL model are compared with experimental results. Finally, s system improvement approach, increasing task engagement, is evaluated using the model.
- Chapter 5, *Hypothesis 2: Moments of Terror*, describes the second human subject experiment that evaluated the ability of the PAL model to predict human performance in responding

to an emergency event after hours of boredom. The effectiveness of restricting external distraction sources is predicted. The quantitative predictions made by the PAL model are compared with experiment results. Finally, s system improvement approach, adding a secondary testing task, is evaluated using the model.

- Chapter 6, *Hypothesis 3: Task Difficulty*, describes the third human subject experiment that evaluated the ability of the PAL model to predict human performance in responding to an emergency event after hours of boredom with different levels of difficulty. Individual differences are investigated. The quantitative predictions made by the PAL model are compared with experiment results. Finally, the impact of changing task processing capability on performance is assessed using the model.
- Chapter 7, *Conclusions*, summarizes the important results in the development and validation of the PAL model. This chapter also demonstrates potential uses for the PAL model by system designers. The generalizability of the model is discussed along with its limitations. Finally, key contributions and potential future works are presented.

2 A Systems View of Boredom: the Boredom Influence Diagram (BID)

In research settings, there is still debate as to the exact definition of boredom. In the late 1920s, boredom was thought to stem from inadequate vascular responses (McDowall and Wells 1927). A decade later, Barmack (1937) defined boredom as a state of internal conflict, caused by inadequate motivation and a desire to remove oneself from a task. O'Hanlon (1981) defined boredom as a psychophysiologic state resulting from prolonged periods of monotonous stimulation. More recently, researchers have generally gravitated to labeling boredom as an affective, and thus subjective, state of low arousal and dissatisfaction caused by a lack of interest in an inadequately stimulating environment (Fisher 1993; Mikulas and Vodanovich 1993; Pattyn et al. 2008).

In his circumplex model of affect, Russell (1980) places boredom roughly halfway between misery and sleepiness. Thackray (1981) reviewed previous studies and concluded that boredom or monotony does not cause stress. Rather, it is the coupling between monotony and a need to maintain high levels of alertness that elicits considerable stress. Hill and Perkins (1985) broke down boredom into a cognitive component of subjective monotony and an affective component of frustration. Focusing more on the underlying mental processes, Eastwood et al. (2012) defined boredom as the aversive state that occurs when one fails to engage attention and participate in satisfying activities.

In the study of optimal experience, boredom is regarded as a mental state resulting from a low challenge level as compared to individual skill level and the lack of intrinsic motivation (Csikszentmihalyi 2014). We could argue that motivation is part of the cognitive component of boredom, because it affects whether an individual perceives the task as boring or interesting. The multidimensional aspect of boredom highlights the fact that boredom is often linked with other physical and cognitive states such as fatigue (Desmond and Hancock 2001) and vigilance (Eastwood et al. 2012), as well as individual traits such as motivation and personality.

Figure 2-1, the Boredom Influence Diagram, represents the multidimensional causes, effects, and interactions of boredom (Cummings et al. 2015). The concepts and interactions shown in this model represent fields of research across many different disciplines. Understanding these concepts and interactions is the foundation of building a dynamic model.

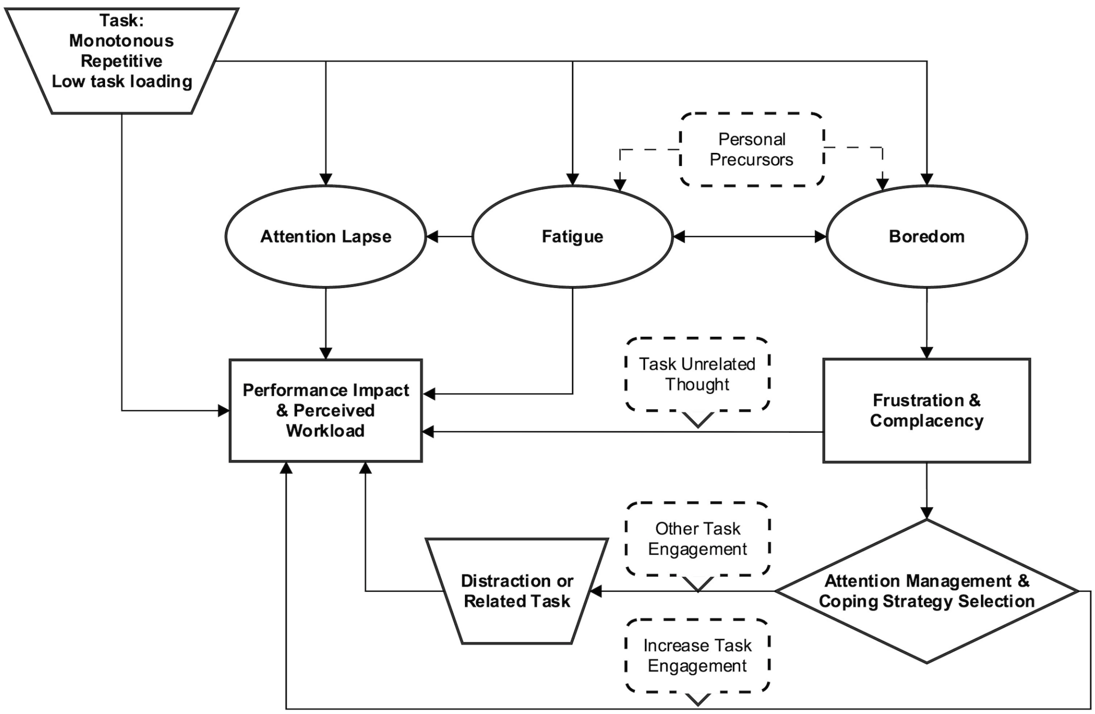


Figure 2-1: Boredom Influence Diagram

The BID framework does not imply cause and effect relationships. Rather, each directional link represents interactions or influences that have been demonstrated or hypothesized in the literature. To begin to understand this diagram, each component will be discussed in the following sections.

2.1 Defining Boring Tasks

As seen in the trapezoid in the upper left corner of Figure 2-1, those tasks likely to be perceived as boring need to be defined, which include monotonous and repetitive tasks in work environments that require constant attention (such as an assembly line task), as well as low task loading scenarios, such as an air traffic controller watching a screen at 2:00 am in the morning, waiting for an aircraft to enter his or her sector. It is important to note the difference between task load (the demands required by the work environment) and workload (the subjective interpretation of task load by an individual), as workload sometimes can be high in monotonous or repetitive tasks, even with low task demand (Warm Parasuraman et al. 2008).

A significant number of previous studies and reviews of boredom in the workplace have focused primarily on environments that include repetitive and monotonous motor tasks such as assembly line production (O'Hanlon 1981; Smith 1981), which is not surprising given the rise of the industrial mass production complex over the first half of the 20th century. More recent studies have shown that boredom occurs in mentally demanding environments that require constant attention (Becker et al. 1991; Dittmar et al. 1993; Prinzel and Freeman 1997; Sawin and Scerbo 1994, 1995; Scerbo et al. 1993). However, there are markedly fewer studies investigating those perceived boring environments where humans are passively monitoring complex systems, waiting for a problem to occur.

As automation has become more prevalent across various work domains, there has been a clear shift away from human manual work on production lines or in direct manual control of vehicles to those environments where humans are supervising automated processes, e.g., in automotive manufacturing plants, robots do the bulk of production line work and in commercial aircraft, pilots spend increasingly amounts of time supervising the autopilot which is actually flying the plane. This increase in automation, however, has not alleviated the boredom associated with these tasks. In many cases, it has exacerbated it, a common phenomenon when more automation is inserted in any system (Bainbridge 1983).

While monitoring a radar or security screening display is very similar to the monotonous vigilance tasks of signal discrimination used in many research settings, monitoring complex automated systems has several different characteristics. Instead of discriminating an event as signal or non-signal repeatedly, people monitoring an automated system have more ambiguous target signals to look for, with typically much longer times between the occurrence of an event. In addition, successful task completion in a complex automated system typically requires much higher situation awareness and problem solving skills.
It should be noted that boredom is a subjective phenomenon, the onset of which is unique to each individual that experiences it. A person's perception of the task at hand may lead to

complacency and cognitive disengagement from the task if the task is perceived to be unimportant or uninteresting. The affective component of boredom reflects a person's emotional perception of the task at hand. These feelings may include frustration, dissatisfaction, or melancholy. For example, boredom may be induced solely as an emotion by asking participants to do nothing (Wilson et al. 2014) or watch uninteresting videos (Merrifield and Danckert 2014).

As proposed in BID (Figure 2-1), three possible behavioral states can occur when a person engages in a task that is monotonous, repetitive, or low task loading: 1) The inability to sustain attention (which is called Attention Lapse), 2) Fatigue, and/or 3) Boredom (represented by ovals in Figure 2-1). These are not mutually exclusive, in that a person could experience one or more of these states simultaneously. Each of these outcomes is discussed in detail in the next sections.

2.2 Attention Lapse

In low task load, highly automated environments, the first likely detectable behavioral outcome for an operator is a lapse in sustained attention, or an ability to maintain vigilance. Vigilance, by definition, is "a state of readiness to detect and respond to certain small changes occurring at random time intervals in the environment" (Mackworth 1957). Typical vigilance tasks, therefore, are naturally repetitive and, at times, could be monotonous and considered to be boring. The vigilance decrement, the decline in performance efficiency over time, is commonly measured in terms of the rate of the correct detection of critical signals and slowed reaction time (Parasuraman 1986).

Monotonous and repetitive tasks have been shown to influence vigilance in a wide range of activities (Parasuraman and Davies 1977), commonly resulting in increases in vigilance decrements, manifested in negative impacts on task performance. The vigilance decrement is commonly measured in terms of missed signals, longer reaction times, and generally poorer performance than can reasonably be expected (Davies and Parasuraman 1982). Vigilance decrements have been demonstrated many times in domains such as aviation (Schroeder et al. 1994; Wiggins 2011), medical monitoring (Weinger and Englund 1990), driving (Thiffault and Bergeron 2003) and rail operations (Haga 1984).

Many studies have tried to explain the mechanism of the vigilance decrement, including mental fatigue (Boksem et al. 2005; Warm Parasuraman et al. 2008), failure in executive control and attention management (Grier et al. 2003), as well as boredom (Scerbo 1998a). None of these factors can fully explain the vigilance decrement. Instead, they are interconnected as illustrated in Figure 2-1. As suggested by Scerbo (1998a), boredom could be the driver for shifting attention away from the primary task, and constantly combating boredom to stay alert could result in stress and fatigue.

2.3 Fatigue

Fatigue can be classified in terms of physiological fatigue or cognitive fatigue, although there is not a crisp defining line between the two, in that the perception of fatigue often drives the interpretation of physical fatigue (Matthews et al. 2012). In a task that involves repetitive gross motor movements, physiological fatigue is common as the body uses its energy reserves. Cognitive fatigue, on the other hand, is generally related to weariness related to depletion of information processing assets (Kahneman 1973; Warm Matthews et al. 2008), reduced motivation (Lee et al. 1991), or stress (Aaronson et al. 1999).

In tasks that are stressful, such as monotonous tasks described previously (Warm Parasuraman et al. 2008), cognitive fatigue will continue to increase as the task duration increases. Fatigue can also be considered as an aggregation of physiological and cognitive fatigue, becoming a sustained feeling of exhaustion that may decrease the ability of a person to conduct physical or mental tasks (Carpenito-Moyet 2006).

There is a distinction between active fatigue and passive fatigue. Active fatigue is derived from continuous and prolonged task-related perceptual-motor adjustment. In contrast, passive fatigue happens in tasks that require system monitoring with either rare or even no overt perceptual-motor response requirements (Desmond and Hancock 2001). In driving, passive fatigue could happen with high levels of vehicle automation, which could reduce driver alertness and increase crash probability (Saxby et al. 2013). Although both passive fatigue and boredom happen under low workload, they reflect different constructs of human cognition. Passive fatigue focuses more on the resource depletion aspect, while boredom reflects the affective state.

2.4 Boredom

While vigilance decrements can be measured and cognitive fatigue induced in people, boredom may be introduced in tasks that do not result in vigilance decrements or cognitive fatigue (Hitchcock et al. 1999; Merrifield and Danckert 2014). Boredom has been described as having two components: a cognitive component and an affective component (Stager et al. 1989). Hill and Perkins (1985) defined the cognitive component as how a person perceives and constructs the task. The affective component comes from the conflict between the inadequate stimuli and the inability to be stimulated in the current environment (Barmack 1939; Fenichel 1951; Hill and Perkins 1985). People are constrained in a boring environment and cannot escape, or they may try to look for new stimuli but fail. The affective state is the coexistence of stimulus-hunger and dissatisfaction, even frustration. In most work environments, such constraints come from production schedules, management policies, and work responsibilities.

Many studies show that vigilance decrement occurs after 20-30 minutes for a task that requires sustained attention (Wickens et al. 2011). It could take a shorter or longer time to observe a vigilance decrement depending on signal modality, signal salience, signal probability, temporal uncertainty, event rate, sleep loss, etc. (Davies and Parasuraman 1982; Loh et al. 2004; Warm et al. 2015). Similarly, boredom can develop as the novelty of a stimulus wears off or the lack of stimuli reaches a satiation point, exacerbated in work environments by the inability to seek new stimuli (Barmack 1939; Scerbo 1998a).

However, there is no consensus and very little research on the temporal aspects of boredom, such as how long it takes to achieve a state of boredom and what conditions or individual differences affect the time at which a state of boredom is achieved. For example, in a study where passive fatigue was introduced by automated driving, task engagement decreased over time, though level of fatigue and boredom were not explicitly measured (Saxby et al. 2013). Moreover, while the vigilance decrement (Warm et al. 1996) is fairly well established across a large cross section of participants and domains, given the subjective nature of boredom, it is not clear if there are any repeatable assumptions that could be made about the onset time and duration of boredom, particularly as these relate to different subpopulations.

The parallel representation in Figure 2-1 of attention lapse, boredom, and fatigue also highlights the experimental difficulties in isolating the effects of one state from the other. Past attempts have been made to assess boredom and people's proneness to boredom, but the experimental research in this area is not as developed as other topics related to attention. Developing an experimental protocol that requires participants to do almost nothing for long periods of time can be much more difficult than designing experiments to test boredom in monotonous task environments that typically last about 20 minutes. Participant recruitment, variation in participants' coping strategies, and measurable data to collect for analysis are just some of the difficulties encountered in investigating low task load studies as opposed to monotonous and repetitive experiments. To effectively study just boredom, we hypothesize that this may not be possible unless the effects of the loss of vigilance and fatigue can be controlled for (either experimentally or statistically). Because it is unlikely that the effects of boredom could be cleanly isolated from the vigilance decrement in the first 30 minutes of a study requiring sustained attention on a task, any experiment that tries to measure boredom in this time period is inherently confounded. This speaks to the need for better boredom assessment strategies.

Due to the disagreement on the definition and underlying theory of boredom, there is no single or widely accepted scale for measuring boredom. Boredom measurement tools have been developed for specific contexts, each with its own advantages and limitations. A few other scales reviewed by Vodanovich (2003) attempt to measure the state of boredom including the Job Boredom, Leisure Boredom, Free Time Boredom, and Sexual Boredom scales. The Multidimensional State Boredom Scale (MSBS) was developed to measure boredom as a state instead of a trait, which includes five factors, namely Disengagement, High Arousal, Low Arousal, Inattention, and Time Perception (Fahlman et al. 2013). State boredom can also be measured using the method of experience sampling in which participants are signaled on a random time schedule to write down information about their momentary situations and psychological states on a self-report questionnaire (Csikszentmihalyi and Larson 1987). Participants can also be asked about how strongly they experienced boredom at a particular moment (Nett et al. 2011). Experience sampling has also been used to investigate Task Unrelated Thoughts (TUTs) (Smallwood et al. 2009).

2.5 Personal Precursors

Since individual traits such as motivation and sleep habits can influence the ability to maintain vigilance and combat fatigue and boredom, they are shown in Figure 2-1 as personal precursors. Given that significant previous research has been devoted to the influence of individual differences on vigilance (Reinerman-Jones et al. 2010; Shaw et al. 2010; Szalma and Matthews 2015; Thackray et al. 1974) and fatigue (Lal and Craig 2001; Matthews et al. 2012; Van Dongen and Belenky 2009), this discussion will focus on the relationship between individual differences and boredom.

2.5.1 Boredom Proneness

Boredom proneness relates to an individual's ability to manage sustained attention tasks (Farmer and Sundberg 1986). Some individuals are more susceptible to boredom than others when facing the same situations that lack external stimuli. Boredom proneness has been positively associated with impatient behavior, distraction, sensation seeking, impulsiveness, and work performance (Dahlen et al. 2005; Kass and Vodanovich 1990; Vodanovich et al. 1991).

The most widely used scale in empirical research to measure boredom proneness is the Boredom Proneness Scale (BPS), which measures boredom as a trait (Farmer and Sundberg 1986). It consists of 28 items (e.g., "It is easy for me to concentrate on my activities"; "Time always seems to be passing slowly"; "I am good at waiting patiently"). The original scale used true-false item format but was later transformed into 7-point Likert scale format. The reliability and factor structure of BPS has been investigated in several studies (Vodanovich 2003). BPS has been used to investigate the relation between boredom and job satisfaction, vigilance reduction, aggressive driving, and many others (Dahlen et al. 2005; Kass et al. 2001; Sawin and Scerbo 1995).

Others have proposed that boredom proneness should be viewed as a multidimensional construct, with external stimulation and internal stimulation as the two primary factors. The external stimulation factor reflects the low level of perceived environmental stimulation and the internal stimulation factor reflects the ability of people to entertain themselves (Vodanovich et al. 2005). External boredom proneness and internal boredom proneness are thought to have different impacts on behavior (Shaw et al. 2010). In one driving study, external boredom proneness was found to contribute to close calls or near misses, while internal boredom proneness predicted reduced adaptive driving anger expression (Dahlen et al. 2005).

There are a few other measures of personal traits related to boredom. The Boredom Susceptibility Scale is a subscale of the Sensation Seeking Scale (Zuckerman 1971). One study that compares the Boredom Proneness Scale and the Boredom Susceptibility Scale shows that they relate to different personality traits and behaviors (Mercer-Lynn et al. 2013). Building on the Boredom Susceptibility Scale and other scales, Hamilton et al. (1984) developed the Boredom Coping scale, which focuses on how individuals restructure their perceptions and participation in potentially boring activities to cope with boredom.

Boredom proneness can play a large role in both success and failure in managing low task load environments in terms of performance. People who are reportedly less prone to boredom have been shown to have better performance in vigilance tasks in terms of the frequency of signal detection as compared to people who score high on the boredom proneness scale (Sawin and Scerbo 1995). In another study, boredom-prone medical and clinical laboratory technologists received lower performance rating from their supervisors (Watt and Hargis 2010). Boredom proneness also correlates with boredom at work and impacts work satisfaction and absenteeism (Kass et al. 2001).

2.5.2 Personality

The widely used personality scales used to investigate boredom-related attributes are different versions of the NEO Personality Inventory and HEXACO. The NEO Personality Inventory measures five factors of personality including Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness (Costa and McCrae 1992). HEXACO uses a six-dimensional structure containing Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience (Ashton and Lee 2007). Extraversion has been associated with boredom proneness (Ahmed 1990), but high levels of conscientiousness have shown the opposite effect (Mkrtchyan et al. 2012). Other personality traits of those people able to more effectively cope with boredom include the ability to spend time alone, high measures of attentional capacity, and low formal diagnostic indices of psychopathology (Hamilton et al. 1984).

The relation of personality and boredom proneness has been examined in a few studies (Culp 2006; Shaw et al. 2010). Another study found that external boredom proneness was negatively associated with Honesty/Humility, Emotionality, and Conscientiousness. Internal boredom proneness was related directly to Extraversion, Conscientiousness, and Openness to Experience (Culp 2006). In one effort attempting to examine the impact of individual differences on vigilance performance more holistically, a factor analysis was conducted based on measures of personality, cognitive–energetic scales, fatigue vulnerability, boredom proneness, sleep quality, cognitive dysfunction, abnormal personality, impulsiveness, cognitive ability, stress states, and coping (Shaw et al. 2010). Four key factors were determined to be cognitive disorganization, heightened experience (defined by unusual experiences, sensation-seeking, and low internal boredom), sleep quality, and impulsivity. In addition, experience, age, intellectual capacity, cultural background, and gender have all been suggested as contributors to the perception of boredom (Fisher 1993; Vodanovich and Kass 1990a). Males tend to exhibit more proneness towards boredom than females (Sundberg et al. 1991), and older people tend to be less susceptible to boredom (Vodanovich and Kass 1990a), although neither of these results are universally found across studies.

2.5.3 Other Factors

Besides personality, experience, age, intellectual capacity, and gender have all been suggested as contributors to individual variability of perceived boredom (Drory 1982; Fisher 1993; Harvey et al. 2010; Thackray et al. 1974; Vodanovich and Kass 1990b).

The frequency of playing video games may provide additional insight as to what kind of person performs well in potentially boring environments. In one experiment looking at the degree of video gaming and performance in a boring low task load Unmanned Aerial Vehicle (UAV) control environment, frequent gamers performed worse that those who were not gamers (Cummings et al. 2013). These same gamers performed well under high workload conditions (Cummings et al. 2010), which raises the question as to how personnel should be selected given potential exposure to both low and high task loads.

A person's interest or motivation in assigned workplace tasks also likely has an impact on an individual's state of boredom (Fisher 1993; Sawin and Scerbo 1995). In one study, individual interest in simple tasks was manipulated by asking participants to set higher goals, resulting in improved performance with reduced boredom (Locke and Bryan 1967). However, given the subjective nature of boredom, individuals will differ in their level of interest in a specific activity, and some can report extreme boredom and others sufficient interest even in an identical environment (Fisher 1993). Boredom has been cited as a direct cause for recruitment and retention issues for the US Air Force's UAV workforce (Cummings 2008), which is problematic since such operators are highly skilled and take years to train.

The effect of boredom on work performance is not uniform for all individuals but rather depends on individual differences (Drory 1982). Identifying who is more prone to boredom will be discussed more fully in a subsequent section on measuring and assessing boredom. It has been shown that high boredom proneness people perform poorly on sustained attention tasks (Malkovsky et al. 2012). In another study examining motivation and boredom, boredom-prone workers felt they were underemployed and received less organizational support, and in somewhat of a self-fulfilling prophecy, received lower performance ratings (Watt and Hargis 2010). This highlights a possible negative motivational feedback loop inherent in these systems.

2.6 Attention Management and Coping Strategy Selection

The last major section of the BID is the entry into the attention management and coping strategy selection phase, represented by a diamond in Figure 2-1. As exemplified by rail train drivers and truck drivers who reported that they listen to music or radio, talk to their co-drivers, eat or snack, and drink caffeine while driving to cope with monotony and boredom (Dunn and Williamson 2011; Oron-Gilad and Shinar 2000), operators will often seek out potentially distracting behaviors simply to stay engaged, although the impact on performance may be ineffective. In a four-hour low task environment of one operator supervising four UAVs, participants spent almost half of the time in a distracted state overall suggesting they were bored (Cummings et al. 2013). As the study progressed, this boredom came at a cost of increased reaction times to system prompts to replan and generate search tasks, as well as text messages asking for information.

Just prior to the decision diamond representing boredom coping strategies, the effects of frustration and complacency are included, which could influence the coping strategy selected by an individual. While the tendency toward complacency could be considered a personal precursor, boring work environments can result in or exacerbate complacency.
Boredom leading to complacency is an established behavioral response in the aviation domain (Wiener and Nagel 1988), which likely leads not only to immediate performance implications but also to long term retention concerns, especially in the presence of increasing automation (Parasuraman and Manzey 2010). In addition, in many studies and surveys, people report that they find working in boring environments frustrating (Bruursema et al. 2011; Fisher 1993; Loukidou et al. 2009; O'Hanlon 1981), which likely leads to not only immediate performance implications but also long term retention concerns. It has been shown that that high boredom-prone individuals perform poorly on measures of sustained attention and show increased symptoms of attention deficit hyperactivity disorder (ADHD) and depression (Malkovsky et al. 2012). As a result, it is proposed that frustration and complacency are responses that will likely affect the coping strategies selected by individuals. Since understanding operator coping mechanisms is of critical importance in system design, how these boredom coping mechanisms can influence performance is also explored.

In terms of coping with a stressful task, one previous study proposed three strategies for coping with a stressful task environment including task-focused coping that attempts to formulate and execute a plan of action to deal with the source of demands directly, emotion-focused coping that attempts to deal with the stressor by changing one's feelings or thoughts about it, and avoidance coping by diverting attention away from the problem (Matthews and Campbell 1998). Other research examining the effects of stress and high workload on human performance proposes that people cope with fatigue and excessive workload by reducing effort and lowering their own performance standards (Hockey 1997). A study on boredom in education identified three coping profiles of students: appraisers that try to change their own perspective of the situation, criticizers that believe they can change the situation by voicing their boredom, and evaders that simply try to avoid the boring setting by doing something else (Daniels et al. 2015).

Given this previous research, it is proposed that when faced with a perceived boring task, resulting behavioral changes can be abstracted into one of three categorical behaviors: 1) task unrelated thought, 2) other task engagement (also known as distraction), and 3) changing task engagement. Task unrelated thought and other task engagement are avoidance coping strategies, which represent passive and active forms of shifting attention. For these states, attention is shifted away from the primary task much like the avoiders and evaders from the previous studies. Our categorization of changing task engagement is a task-focused coping strategy, in which attention is allocated towards the primary task. An emotion-focused coping strategy is not included here because boredom itself is an affective state. These three coping strategies are discussed in more detail next.

2.6.1 Task Unrelated Thoughts

One way for operators to cope with boredom and associated frustration and complacency is through Task Unrelated Thought (TUT), also known as mind wandering, stimulus-independent thought, and daydreaming, which occurs when one's mind drifts "from a task toward unrelated inner thoughts, fantasies, feelings, and other musings" (Smallwood and Schooler 2006). Daydreaming and thinking were frequently reported as strategies used to cope with boredom in life (Fisher 1993; Harris 2000). In a study of airline pilots serving in a monitoring role, it was observed that pilots devoted 43% of their available monitoring time to TUT, often when they felt their performance would not conspicuously suffer (Casner and Schooler 2015). The basic implication of TUT is that a person may be physically present in a control environment, but is unable to remain cognizant of the control task at hand. This disengagement is the resulting effect of an endogenously generated distraction, created to cognitively engage the individual and limit the negative affect felt during low task loading.

TUT and self-generated thought are spontaneous processes and the default state of the individual. Based on brain imaging results, neuroscience research has found that the brain is more active at rest than in a range of explicit tasks, possibly because the brain is engaging in self-generated thought (Morcom and Fletcher 2007). Several studies proposed that TUT reflects a failure in executive control (McVay and Kane 2010; Thomson et al. 2015). Instead of devoting attention to TUTs by choice, individuals need to execute explicit control to sustain active goal maintenance and to prevent TUTs. TUTs are considered as spontaneous processes (Christoff et al. 2004).

While TUT can occur in high workload environments, it is generally associated with under stimulating, low task load environments, and has recently been shown to be pervasive across all aspects of life (Killingsworth and Gilbert 2010). TUT consumes attentional resources and reduces the attention devoted to the primary task (Smallwood and Schooler 2006).

In one boring search task, a majority of participants exhibited task disengagement in the form of non-task-related mental activity (Pattyn et al. 2008). TUT has also been found to increase across the duration of a vigilance task, accompanied by a decrease in detection accuracy (Cunningham et al. 2000). Higher levels of automation allow for more TUT as shown in a flight automation study (Casner and Schooler 2014). Thus, the performance impact of TUT on a particular task can be seen as negative when considering missed signals and increased reaction times, as individuals seem to be incapable of engaging in TUT and the task concurrently. It is important to note that TUT is a passive form of coping, in that one is usually not engaged in any physical activity or conversation.

2.6.2 Other Task Engagement

While TUT represents a passive form of task disengagement in perceived boring environments, engaging in tasks other than the primary task represents a more active form of task disengagement. People may seek stimulation intentionally from sources other than the primary task when they feel bored or they may be easily distracted by external activities in the environment. The previous examples of the Northwest pilots working on their laptops, causing them to overfly Minneapolis, and the air traffic controller watching a movie in the early morning hours are examples of such occurrences.

While such events can be seen as distractions, it is important to make the subtle distinction as to the source of the distraction. In the interruption recovery literature (John et al. 2005; Scott et al. 2006), distractions are generally seen as exogenous events that cause an operator to shift attention from a primary to an intruding task. In low task load, boring environments, operators may seek stimulation from a possibly unrelated source, in effect seeking an interruption, and so in this context, the source of such distractions is endogenous and intentional.

As seen in Figure 2-1, the two basic types of 'other tasks' are labeled as Distraction or Related tasks. For example, in many process control plants, operators will often complete training modules during low task loading, so arguably they are somewhat distracted, but by a task that is related to their current job. However, operators in such environments also read magazines or newspapers, which is an unrelated task. How related versus unrelated distractions impact performance is an area that has received very little attention in research settings.

It is not clear whether such distractions in low task load environments always result in poor performance. For example, although talking on one's cell phone while driving has been repeatedly shown to lead to driver distraction in high workload settings (Patten et al. 2004), little research has been done to examine possible positive benefits, such as when driving during long stretches of highway, particularly at night. Military truck drivers have reported cell phone use relieves some monotony experienced during long drives (Oron-Gilad et al. 2008) and thus a possible positive relationship between some distraction and relieving the negative impacts of boredom.

So while it is possible that distraction could reduce boredom during a task, this likely only leads to positive performance benefits when the task requires low levels of attention, such as in monitoring an automated system for an alert. When the task requires more substantial engagement of attention such as what is needed to complete monotonous or repetitive tasks, distraction will not likely relieve boredom (Fisher 1998).

In terms or mitigating the negative affects of boredom, it has been proposed that a secondary task can be strategically embedded in the primary task setting in order to decrease boredom and increase capability and concentration, which would ultimately increase performance and safety (Atchley and Chan 2011). In one driving study, it was found that introduction of a concurrent verbal free association task improved lane-keeping performance and lowered steering wheel deviations in some conditions during prolonged driving (Atchley and Chan 2011). In another diving study, Oron-Gilad et al. (2008) compared answering trivia questions, a choice reaction time task, a working memory task, and listening to music as secondary tasks to help drivers stay alert. It was found that the trivia task prevented driving performance deterioration and increased alertness, while the working memory task was detrimental to driving.

However, this strategy of introducing a secondary task must be used with caution. The attention requirements of the primary and secondary tasks must be carefully evaluated to avoid any negative impact. If the secondary task requires little cognitive effort, it could result in positive effects such as reducing boredom. However, in more complex cognitive engagement tasks, such embedded secondary tasking may cause overload and result in decreased performance.

2.6.3 Changing Task Engagement

The last possible coping strategy category, changing task engagement, is another area that has received little attention in the literature. High level of task engagement is characterized by high energetic arousal, task interest, success motivation, and concentration (Matthews Warm Reinerman-Jones et al. 2010). Behaviorally, operators become aware that the task load is low, and interact with the system to change their task load to stave off any negative effects of boredom. For example, night watchmen in charge of monitoring several cameras may constantly manually pan and zoom in order to stay alert. Changing the primary task engagement includes accessing task-related imagination, refocusing attention on the task, and increasing or changing the complexity of the task.

Task-related imagination turns the primary task into a game or mental cinema, which may increase the degree of intrinsic interest in the task and reduce boredom (Eastwood et al. 2012). In contrast to TUT which diverts attention to unrelated thoughts, the imagination is engaged to consider the task at hand. Further, this may improve task performance by facilitating absorption thereby attenuating the experience of attention failure, effort, and boredom, which would then promote successful engagement with the current task (Csikszentmihalyi 1978; Eastwood et al. 2012).

A related technique called gamification has been used in education settings by including game-like elements to engage bored students (Barata et al. 2013). Some have proposed applying gamification for driving (Schroeter et al. 2014), but its utility remains untested in this setting. In one process control study, allowing operators to play a nuclear power optimization game in parallel with a low workload primary task did not improve performance but neither did it degrade it (Thornburg et al. 2011).

Another method to change task engagement is to simply notify or alert the operator periodically so that he/she could refocus attention to the task. One study has shown that such a strategy can be useful for operators prone to boredom and distracted for a considerable amount of time (Mkrtchyan et al. 2012). Others have suggested that using biofeedback (such as using electroencephalogram (EEG) to monitor physiological processes) and displaying this information to an operator in real-time can potentially alert an operator to refocus. One study showed that TUTs can be reflected in EEG power band ratios in the intervals immediately preceding and following the subject's report of a TUT (Cunningham et al. 2000). It has been proposed that such feedback could be beneficial in stimulating cognitive activity and reducing boredom during monitoring tasks (Alves and Kelsey 2010; Frederick-Recascino and Hilscher 2001).

The third form of changing task engagement is to modify the complexity or requirement of the primary task. This could be initiated either by human activity or by the system. In the driving scenario for example, drivers may try to maintain some level of arousal by using adjustments in speed to change the task difficulty (Fuller 2005). They may increase speed as a response to under stimulation, especially young male drivers (Heslop et al. 2009). In a nuclear power plant example, human operators have the option to examine individual systems components in more detail from the control console. UAV operators can elect to bring up new displays through various menus to consider new sources of information.

However, the system could be designed to increase task requirements in order to increase operator engagement. In one air traffic control monitoring study, task engagement was increased by requiring the controller to click on each aircraft as it entered the airspace, which mitigated the vigilance decrement after the operators were sufficiently trained for the task (Pop et al. 2012). Task engagement can also be adjusted dynamically by varying task difficulty according to the measurement of operator status as revealed through functional near-infrared spectroscopy (fNIRS) signals. Afergan et al. (2014) used such signals to increase operator awareness and reduce errors. In video game design, increasing task engagement by adjusting difficulty levels and skill requirements over time is a common method to avoid boredom as well.

Changing or increasing task engagement may have positive or negative influence. Obviously, seeking stimulation by speeding up raises safety concerns during driving. On the other hand, some cases in European cities show that creating a shared space between cars, bikers, and pedestrians on the road could surprisingly increase safety (Hamilton-Baillie 2008). Although the underlying mechanism is not entirely clear, one possible reason is that shared space forces drivers to devote more attention and effort into driving, reducing the tendency to speed when they feel under stimulated.

Whether increasing task engagement could improve performance also depends on the level of additional cognitive demands placed on the operator. In the air traffic control monitoring study mentioned earlier (Pop et al. 2012), engaging the operator by requiring the controller to click on each aircraft as it entered the airspace alleviated the vigilance decrement after practice. However, when the engagement task required increasing attention, the vigilance decrement could not be eliminated because the engagement task was competing with the primary task, resulting in operator cognitive overload.

Another concern that arises when operators elect to increase their primary task engagement by interacting more with their system is whether that system is robust to the increased interactions. For example, in one study where participants performed a decentralized multiple UAV control task, operators that interacted too frequently with a system harnessing optimization algorithms could actually drive the system to a sub-optimal state (Cummings et al. 2012). So in some cases it is possible for an operator to attempt to alleviate boredom by interacting with the system, which could ultimately result in degraded system performance.

It should be noted that all three of these coping strategies (engaging task-related imagination, refocusing attention on the task, and increasing or changing the complexity of the task) could all be present for a single operator over the course of a single shift in a low task or monotonous environment. More research is needed to both observe if and how people vary these strategies to combat the negative effects of boredom and how such application of strategies can either improve or degrade overall systems performance.

2.7 Performance Impact and Perceived Workload

The final block in the BID depicted in Figure 2-1 is that of Performance Impact and Perceived Workload, which is clearly a critical outcome. Regardless of the task type or the coping strategy, the BID in Figure 2-1 demonstrates that lapses in attention, fatigue, and boredom can occur in parallel, ultimately influencing system performance and operator workload. It should be noted that these attentional lapses could be both episodic, as well as persistent states.

2.7.1 Performance Impact

The influence of fatigue on performance is well documented (Krueger 1989). For example, operators of UAVs in long duration missions will commonly rate their feeling of fatigue to be very high (Chappelle et al. 2014). Such affective states, particularly negative, can greatly influence human performance (Norman 2004), and in the case of UAV operations, cognitive fatigue has been shown to result in slower responsiveness and reduced task performance (Thompson et al. 2006).

Moreover, loss of vigilance can cause delayed response, missed signals, and increased false alarms. Such results parallel similar results in vigilance studies, where reaction times, false alarms, and stimuli missed increase over time, and such changes in performance are likely an indicator for mental fatigue, which could be influenced by boredom (Azarnoosh et al. 2012; Ballard 1996; Scerbo 1998b). Boredom and individual coping strategies affect performance indirectly by changing attention allocation. Smallwood et al. (2004) suggest that although high levels of TUT can happen together with increased errors in sustained attention tasks, TUT is not the direct cause of performance decrement. In general, fatigue, boredom, and loss of vigilance result in a decrease in task performance.

However, one problem remains in the measure of such performance change. By definition, boring work without monotonous tasking, like that seen in process control plants where operators monitor a plant for several hours without ever touching a control device, have little stimulation and thus few observable events. For manual driving and semi-automated driving, performance can be monitored based on steering behavior, speed control, and lane keeping. In highly automated driving, these variables no longer provide much information as the automation is in control (Merat et al. 2012; Saxby et al. 2013).

What is not explicitly represented in Figure 2-1 in terms of performance impact but likely is significant is the temporal factor. For example, if some operators tend to cope with boredom by increasing their own workload either through introducing endogenous tasking or distractions, are they then more prone to fatigue over time, which could influence complacency and/or frustration? How can we better model the temporal influence of boredom and passive fatigue in low task environments? These questions are areas for future investigation, as there is little in the current literature to address the temporal aspects of boredom.

2.7.2 Perceived Workload

In addition to performance, perceived workload is also affected by task demand, fatigue, and boredom. It has been shown that workload is the highest under active fatigue with difficult manual driving, and lowest under passive fatigue with automated driving (Saxby et al. 2013). In several studies with vigilance tasks requiring participants to discriminate between signals, monotonous vigilance tasks are often rated as high in workload and stressful (Finomore et al. 2013; Warm et al. 2015; Warm Parasuraman et al. 2008). Thus workload is influenced by task demand, which can be correlated with monotony of the task and the degree of automation. In addition, subjective workload is also found to follow an S curve with unitary increases in working memory load (Estes 2015). However, workload and boredom can be manipulated independently. In a vigilance task, cueing the arrival of a signal can decrease workload while keeping the boredom level of task unchanged (Hitchcock et al. 1999).

Boredom assesses both the external environment stimuli and internal personality traits, while workload measures are about an individual's ability to cope with the task requirements. Thus while subjective workload rating scales such as the NASA-TLX have been validated in a number of high workload studies (Hart 2006), it is not clear whether such workload scales can accurately capture the influences of boredom. More importantly, although boredom often happens in low workload environments, it can also occur in high workload environment where the task is monotonous or repetitive (Warm Parasuraman et al. 2008).

This introduction and discussion of the BID is useful in understanding the influence, interactions, and performance implications of working in a boring low task loading and/or monotonous environment. However, one critical component to making such a framework useful is identifying those people, processes, and coping strategies that lead to better outcomes in such an environment by assessing and measuring the impact of these different aspects. The next section will outline commonly used assessment strategies and measures for operators working in boring low task load and/or monotonous environments, with an emphasis on those environments where automation plays a significant role.

2.8 Modeling Boredom

Boredom is a universal state that can happen in many different settings. Recently, there are a few models developed using machine learning methods to detect boredom based on physiological signals and behavioral patterns. These models are used to detect boredom in leisure time and entertainment. For human automation interaction, there are no dynamic models to capture boredom or the influence of boredom. There are some theoretical models to explain vigilance decrement, which are introduced in this section.

2.8.1 Detecting Boredom

Several studies have attempted to build classification models for boredom using machine learning methods, with reported good accuracy. In one study where participants performed anagram-solving tasks while playing Pong, their emotional states were assessed using a self-report questionnaire and physiological signals. Three intensity levels of boredom (high, medium, low) were then classified with an accuracy of 84.23% based on signals including electrocardiogram, bio-impedance, electromyogram (from the corrugator, zygomaticus, and upper trapezius muscles), electrodermal activity, peripheral temperature, blood volume pulse, and heart sound (Rani et al. 2006). In another study where participants played 3D video games, using moment-based features of electrocardiogram (ECG) and Galvanic Skin Response, binary classification accuracy was reported to be 94.17% for the states of bored or not bored (Giakoumis et al. 2011). Unfortunately because these studies are very specific to the test beds used and contain low numbers of individual participants with different feature sets, it is difficult to compare the studies and draw definitive conclusions. Moreover, in these studies, people were engaged in an activity quite different from scenarios in which people do nothing.

A few other studies have attempted to detect boredom from behavior patterns. One study classified mobile phone users into low and high boredom proneness groups with over 80% accuracy (Matic et al. 2015). Key predictors are the number of received social network notifications, frequency of notifications, and changes in screen status. Another study attempted to detect the state of boredom based on correlations between mobile phone usage patterns and subjective ratings of boredom level from real-time sampling (Pielot et al. 2015). Some of the features included recency of communication activity, intensity of recent usage, general usage intensity, context or time of the day, and demographics. Furthermore, in a second in situ study, it was found that users were more likely to engage with suggested content on their phones when bored. Using a different approach, Kapoor et al. (2015) used a Hidden Semi-Markov Model to predict the gaps between user consumption activities. If users get bored with a particular item they were engaging with before, they would move to a different item of interest, and return to the original item of interest after a gap period.

A major limitation to these models is that they do not explicitly capture boredom at work, or as a function of a perceived boring work environment. In addition, these models focus only on the state of boredom while ignoring other factors such as fatigue, attention and the impact on performance. However, the usage of physiological signals and behavior patterns can inspire the investigation of boredom when interacting with automation.

2.8.2 Modeling the Influence of Boredom

There are several theories that attempt to explain vigilance decrements in performing vigilance tasks. While interacting with automation with low task demand is different from the repetitive vigilance tasks, investigating these theories is still helpful. In both types of tasks, sustained attention is required to achieve good performance. One theory explains the decrease of performance in vigilance tasks based on the depletion of information processing resources, which is also called the overload theory (Warm et al. 1996), reinforced by another study that showed that vigilance tasks are taxing and effortful (Warm Parasuraman et al. 2008). The depletion of information processing resources results from two factors: the task demand and the time on task (Caggiano and Parasuraman 2004).

Another theory is the underload or mindlessness hypothesis (Manly et al. 1999). Based on this theory, a decrease in task performance in vigilance tasks happens because people withdraw attention and respond to tasks automatically and mindlessly. There are debates around these two theories. Recent studies comparing the two theories show that the overload theory is more convincing (Grier et al. 2003; Helton and Warm 2008).

However, none of these theories are sufficient to explain the performance decrement in boring task environments with little tasking. In vigilance tasks, people frequently need to decide whether a stimuli is a signal or not (e.g., a sonar operator looking for a signal). When interacting with a highly automated system in which a task needs human interaction only occasionally, people have almost nothing to do but monitor the system. This does not result in overload. Cognitive resources could still deplete from time on task, but very slowly, with a reduction of attention on primary tasks. This is more of a failure in effortful attention rather than mindlessness (Grier et al. 2003).

A recent paper by Thomson et al. (2015) brings these two theories together and proposes a resource-control theory as shown in Figure 2-2. In this theory, the total amount of attentional resources remains constant, and the amount of resources needed for the primary task remains constant. However, executive control decreases over time-on-task, resulting in a disproportionate amount of resources being devoted to mind wandering and not enough resources being devoted to the primary task, resulting in performance costs. One limitation of this theory is the assumption of constant resource requirement. In real world tasks, task demand is rarely constant. Especially when an emergency event occurs, there is a quick shift from low task load and high task load. However, it is difficult to model such dynamics in a static model.

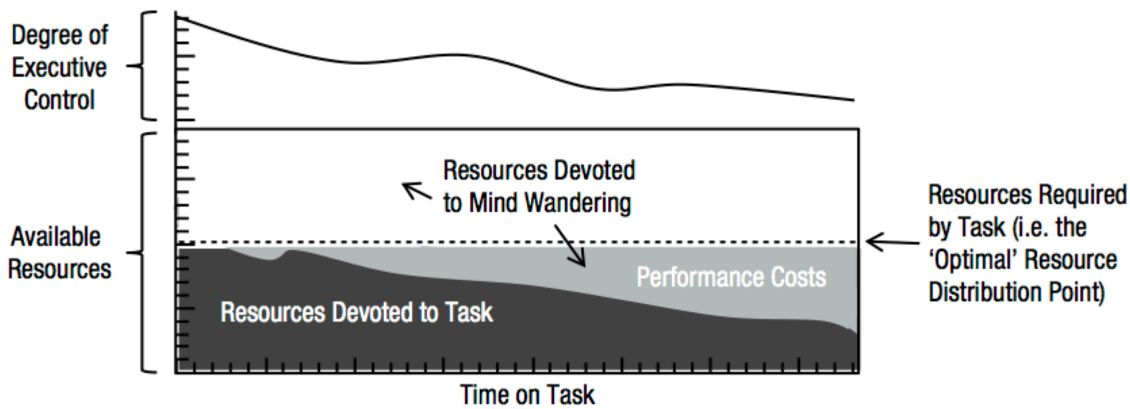


Figure 2-2: The Resource-Control Account of Sustained Attention(Thomson et al. 2015)

One key element of the resource-control theory is the reduction of executive control over time. One theory to explain this is the resource theory. One program of laboratory studies suggests that self-control depends on a limited resource of executive control, which is depleted with acts of self-control and restored with rest and positive affect (Baumeister 2002). This limited resource, called executive control resource, is allocated to a variety of specific processes including information processing, attention allocation, self-regulation, decision-making, etc. (Gross and Grossman 2010). Many studies found that acts of self-control at Time 1 reduce performance on subsequent, seemingly unrelated self-control tasks at Time 2 (Baumeister et al. 1998; Inzlicht and Schmeichel 2012; Muraven et al. 1998; Schmeichel 2007). In the model of boredom developed by Hill (1985), boredom arises when the stimuli are monotonous or inadequate, and the person cannot find additional or alternative stimulation. The requirement to stay focused on the task despite its repetitiveness and boring nature depletes executive control resource, which is also called "ego depletion" (Baumeister 2002; Baumeister et al. 1998). When the executive control resource is depleted, people have difficulty in sustaining attention.

There are other theories explaining the decrease of executive control resource other than the resource theory. Several studies show that motivation plays an important role in executive control. One study suggests that exerting self-control at Time 1 causes temporary shifts in both motivation and attention that undermine self-control at Time 2 (Inzlicht and Schmeichel 2012). Another study shows that depleted individuals may compensate for their lack of self-control resources when sufficiently motivated (Muraven and Slessareva 2003). A more recent paper completely discards the resource account, and suggests that regulatory failures reflect the motivated switching of task priorities as people strive to strike an optimal balance between engaging cognitive labor to pursue 'have-to' goals and the pursuit of 'want-to' goals (Inzlicht et al. 2014).

2.9 Summary

While the issues of boredom in the workplace in general, and more specifically in highly automated environments, are known to researchers and practitioners, they have generally not been as well researched as in other areas such as vigilance and the effect of high workload on performance. This section presented a framework by which to organize the various facets of boredom, particularly in supervisory control settings.

The research that led to this framework highlighted many gaps. First, much previous research focused on monotonous and repetitive tasks, while boredom in low task loading environments is less understood. Even traditional vigilance tasks have repetitive features. Second, there is no framework that systematically investigates boredom and its influence in supervisory control settings. Third, while there are theoretical models to explain the vigilance decrement and the decline of attention, dynamic models are not available to capture the temporal changes. Because of the move towards more automated systems in the future, a better understanding is needed to enable intervention and mitigation of possible negative impacts. These research gaps highlight the need to develop a systematic model that can capture the dynamic change of task load, workload, as well as the influence of boredom on attention and performance. Based on the concepts and interactions presented in BID, it is possible to build such a model using system dynamics modeling. This is introduced in Chapter 3.

3 Modeling and Simulation

This chapter describes the modeling process that created the Performance and Attention with Low-task-loading (PAL) Model, a System Dynamics (SD) model of human-automation interaction in long duration, low task load scenarios. These are task environments with high levels of automation and humans are passively monitoring complex systems most of the time. The chapter begins by describing the field of System Dynamics and why it is appropriate for modeling human automation interaction in low task load scenario. The model building process is then described, with five subsystems discussed in detail. Three dynamic hypotheses are proposed regarding human attention and performance. Next, the results of model structure tests are presented. The chapter concludes by summarizing the benefits and limitations of the model.

3.1 System Dynamics Modeling

System dynamics (SD) is an approach to understanding the behavior of complex systems over continuous time, where the change of the system happens at small time steps (Sterman 2000). SD models deal with internal feedback loops and time delays that affect the behavior of the entire system. The main components of SD models are stocks, flows, and causal loops. SD models have been used in many large and small scale systems with social elements including management, economics, logistics, education, and epidemics (Sterman 2000). Human behavior such as bounded rationality (Morecroft 1985) can be captured using SD models. More relevant to human-automation interaction, SD models have been used to model disasters resulting from the accumulation of routine interruptions to existing plans and procedures (Rudolph and Repenning 2002), procedure rework in space shuttle mission control (Owens et al. 2011), worker burnout (Homer 1985), human problem-solving under time-pressure in action-oriented environments (Rudolph et al. 2009), and real-time human-automation collaborative scheduling (Clare 2013). These models focus on high workload domains, and no previous efforts have been devoted to model issues result from low task loading.

While other simulation modeling techniques, such as Discrete Event Simulation (DES) and Agent-Based Modeling (ABM), have been successfully applied to modeling human supervisory control (Gao et al. 2014; Nehme 2009; Ryan 2014), there are a few reasons to use SD to model human attention and performance in low task load automated environments. This approach is well suited to modeling continuous processes, systems where behavior changes in a non-linear fashion, and systems where extensive feedback occurs within the system. In addition, SD is very useful for the modeling of both quantitative and qualitative aspects of a system. There can be many qualitative aspects of a system that are difficult to quantify, but are important for the behavior of the system. In SD models, these aspects are modeled based on reasonable assumptions and expert opinions (Sweetser 1999). In order to model human performance under low task load, constructs such as workload, fatigue, boredom proneness, and stress all need to be included, which are sometimes difficult to measure and quantify. This makes SD an ideal tool for such purpose. Lastly, SD takes a system point of view and considers the interaction among different factors and processes. The use of causal links and feedback loops could generate system behaviors that are not anticipated intuitively. As in the BID, which was discussed in Chapter 2, human performance under low task load is closely related to several interconnected constructs. SD is well suited for modeling such systems.

While SD models are useful for clarifying the complexities of system behavior, the use of average flow rates and the necessity to aggregate entities are recognized as significant limitations (Brailsford et al. 2010). Usually, SD models are deterministic instead of stochastic. The value of a flow rate is fully determined by model structure, ignoring any randomness in behavior. This could introduce bias into the model if stochastic behavior is important. In such cases, one can replace constant values in the SD model with probability distributions or use other stochastic modeling approaches. Also, system dynamics models operate at a much more aggregate level by concentrating on the rates of change of populations of entities (Morecroft and Robinson 2005). While the goal of many SD studies is to model average behaviors of humans, accuracy of such model predictions will be influenced if individual differences are not included. Thus SD models are not particularly useful for modeling individuals but they are useful for representing the impact of individual characteristics on the system as a whole. Another critique of SD models is the tendency to over fit the model because of the excessive degrees of freedom provided by the input variables. To reduce the impact of these limitations, rigorous modeling procedures and model testing methods should be employed during the entire model developing and testing stages.

The SD modeling procedure has five major phases (Sterman 2000). First, in the problem articulation stage, the overall problem that the model is trying to represent is identified, along with key variables to be captured and the boundaries of the model. This part of work was described in Chapter 2. In the second phase, dynamic hypotheses are developed. A dynamic hypothesis is a theory that explains the behavior of the system as an endogenous consequence of the feedback structure of the holistic system (Sterman 2000). It guides the modeling effort and is continuously tested and refined throughout the model building and testing process. Section 3.2 describes the three dynamic hypotheses in this research. In the third stage, the dynamic hypotheses are mapped into causal loops and stocks and flows, as described in Section 3.3. The fourth stage, testing the model, includes model structure testing and comparison of model outputs to experimental data sets. The fifth stage is policy design and evaluation, including evaluating the ability of this model to predict performance under new circumstances. Since three task scenarios were used to test the three hypotheses, Chapters 4, 5 and 6 each focus on one dynamic hypothesis, and present the modeling testing and policy evaluation for each task scenario.

3.2 Dynamic Hypotheses

Based on literature as discussed in Chapter 2, three dynamic hypotheses were developed. A dynamic hypothesis is a theory that explains the problematic behavior in a system (Sterman 2000). In low task load environments, the behavior of concern is the decrease in human performance over time. To develop strategies to mitigate the negative impacts, the key is to understand the underlying mechanisms that lead to the decrease in performance. The three dynamic hypotheses explain the mechanisms starting from attention changes, to performance impact, and finally the impact of task difficulty in low task load automated environments.

In Chapter 2, the relation between attention and performance was discussed. The failure in attention management is a key reason for the decreased performance in low task load environments. Research shows that vigilance begins to decline after 20-30 minutes for a task that requires sustained attention (Wickens et al. 2011). To this end, the first dynamic hypothesis deals with the change of attention in low task load environments, as demonstrated in Figure 3-1. It was hypothesized that low task load environments result in reduction in executive control, which causes failure in attention management (Matthews Warm Reinerman et al. 2010; Thomson et al. 2015). Hypothesis 1 is the foundation for the other two hypotheses, which further investigate the impact of such attention changes on performance.

Hypothesis 1: On average, individuals reduce their attention on primary task under low task load.

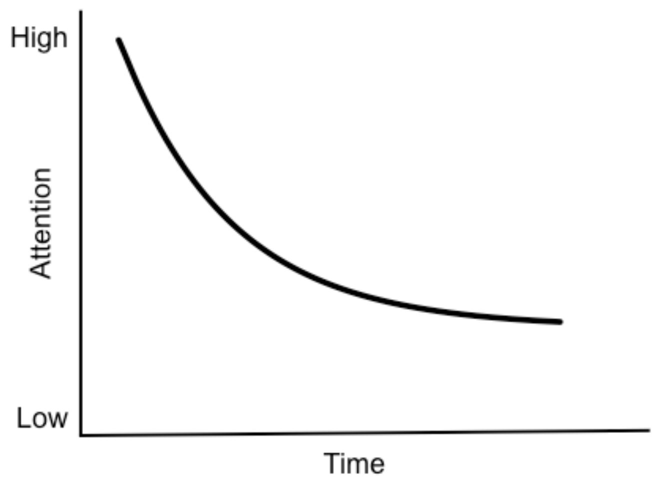


Figure 3-1: Decrease of Attention over Time

The second hypothesis deals with the impact of reduced attention on performance. According to the Yerkes-Dodson Law, performance is impaired in both high and low arousal or stress levels (Teigen 1994). While automation can handle most of the tasks in an automated supervisory control environment, an unexpected event might need human intervention either due to the limited capability of automation or automation failure. In such cases, a human operator's stress level quickly increases. When this human operator has a low attention and executive control resource level, this can further introduce excessive stress. Anxiety has adverse effects on processing efficiency by impairing three executive functions—mental set shifting ("Shifting"), information updating and monitoring ("Updating"), and inhibition of task-irrelevant stimuli ("Inhibition") (Eysenck and Calvo 1992; Eysenck et al. 2007). Instead of reaching optimal performance in dealing with the unexpected event, performance can be impaired with excessive stress and anxiety, as shown in the comparison of A and B in Figure 3-2.

Hypothesis 2: The reduction of executive control resource and attention on a primary task under low task load leads to higher stress when unexpected tasks happen, and worse human performance in dealing with unexpected tasks.

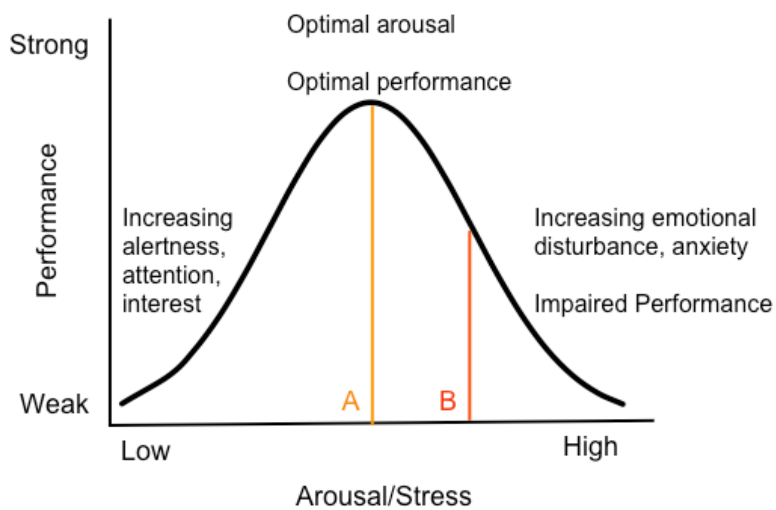


Figure 3-2: Yerkes-Dodson Law

While the inverted-U curve of the Yerkes-Dodson Law is widely used, the original version based on the actual Yerkes-Dodson findings also takes into account the differences of simple tasks, such as when the task involves focused attention on a restricted range of cues, and more complex or challenging tasks, such as in divided attention, multitasking, and working memory tasks (Diamond et al. 2007). With the high level of stress resulted from unexpected event, the performance of difficult tasks is impacted more than simple tasks, as shown in Figure 3-3. This leads to the third dynamic hypothesis:

Hypothesis 3: With reduced executive control resource and attention on the primary task, human performance on unexpected tasks in low task load supervisory control settings is worse with difficult tasks as compared to easy tasks.

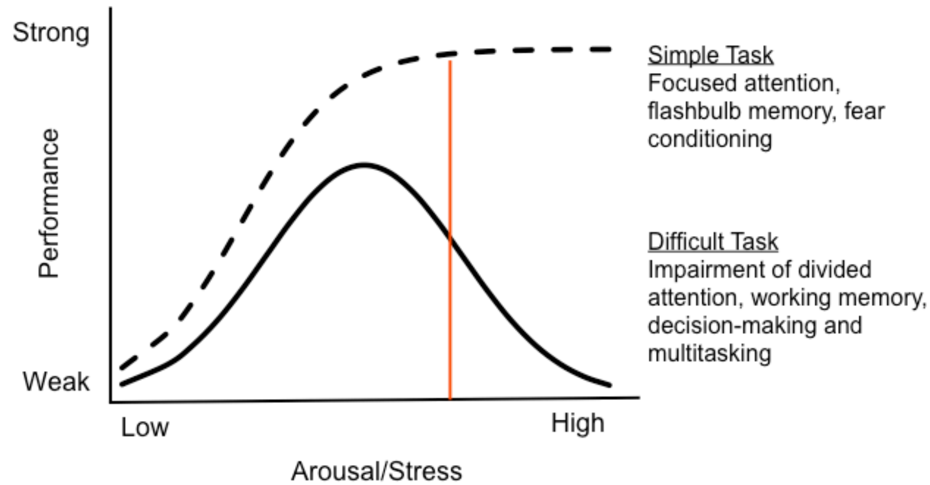


Figure 3-3: Yerkes-Dodson Curve with Different Task Difficulty

These three dynamic hypotheses are tested using three experiment datasets, which are presented in Chapter 4 to Chapter 6.

3.3 Model Implementation

This section describes the process that created the PAL Model of human attention and performance in automated environments with low task loading. This model has been developed using SD modeling techniques, drawing from the results of previous human-in-the-loop experiments and supporting literature in human factors, cognitive science and psychology. The model was built through several iterations to refine the model structure, as described in Appendix B. The final model is described in this section. Factors affecting operator performance are captured in the model as five subsystems: 1) task characteristics, task processing and performance; 2) stress; 3) workload; 4) executive control; and 5) attention management.

The PAL Model simulates the operations of human-automation interaction with low task loading. It could be extended to high task loading environments. However, this is not the scope of the research and is left for future research. In these automated systems, human operators only need to interact with the system occasionally to update the task schedule or respond to rare events while the automation takes care of the routine work. Often operators are required to monitor the status of the system and stay vigilant. In addition, a shift from low task load to high task load can happen during such a mission. By modeling the interaction of task load, stress, and executive control, this model provides metrics on human performance and attention throughout the mission. The model implements a set of equations that are calculated at discrete time steps using the Vensim® simulation software package.

3.3.1 Model Overview

A simplified diagram of the model is shown in Figure 3-4, which depicts five modules and six major feedback loops. The five modules are:

- *Task Characteristic, Processing and Performance*. This module is represented by two blocks in Figure 3-4. The block of *Task Characteristic* models the arrival process of tasks and the time within which these tasks need to be finished. The block of *Task Processing and Performance* models the speed of task processing, the number of pending tasks, and the number of completed tasks.
- *Stress*. This module models the stress level of an operator, which is affected by task requirements, executive control and attention management. Stress level also influences task processing and attention management.
- *Workload*. This module models the workload level of an operator as affected by the task load. When the task load is high, the perceived workload is high.
- *Executive Control*. This module models the depletion and recovery of executive control resources. Executive control resources are depleted when the operator feels bored from the low task load. They are recovered when the operator engages in activities unrelated to the task.
- *Attention Management*. This module models the dynamic attention allocation process of an operator. Attention is allocated to or shifted away from the task as a result of task demand and executive control.

The five modules are connected via six feedback loops: Yerkes-Dodson Loops, Ego-Depletion Recovery Loop, Refocus Loop, Drained from Boredom Loop, Attention Control Loop, and Increase Task Engagement Loop. An overview discussion of the loops is presented below.

The Yerkes-Dodson Loops, shown in black arrows in Figure 3-4, describe the relation between task requirement, stress and task performance. Yerkes and Dodson (1908) found that both low and high levels of arousal yield low performance. Best performance is achieved when the person is neither under nor over loaded. While much research focuses on the right half of the inverted-U curve, understanding the impact of low task load on performance is becoming more and more important with the development of automated systems. In the system dynamics model, the Yerkes-Dodson effect is represented by two separate loops, each corresponding to either positive or negative effect of stress.

The Ego-Depletion Recovery Loop is highlighted in blue arrows. Ego depletion refers to the idea that self-control or willpower is a limited pool of mental resources (Baumeister 2002). Self-control is typically impaired when this mental resource is low, which can be recovered with a break (Baumeister et al. 1998). When people get bored with low task loading, some distraction or a break can sometimes help them refocus. This phenomenon is captured by this loop.

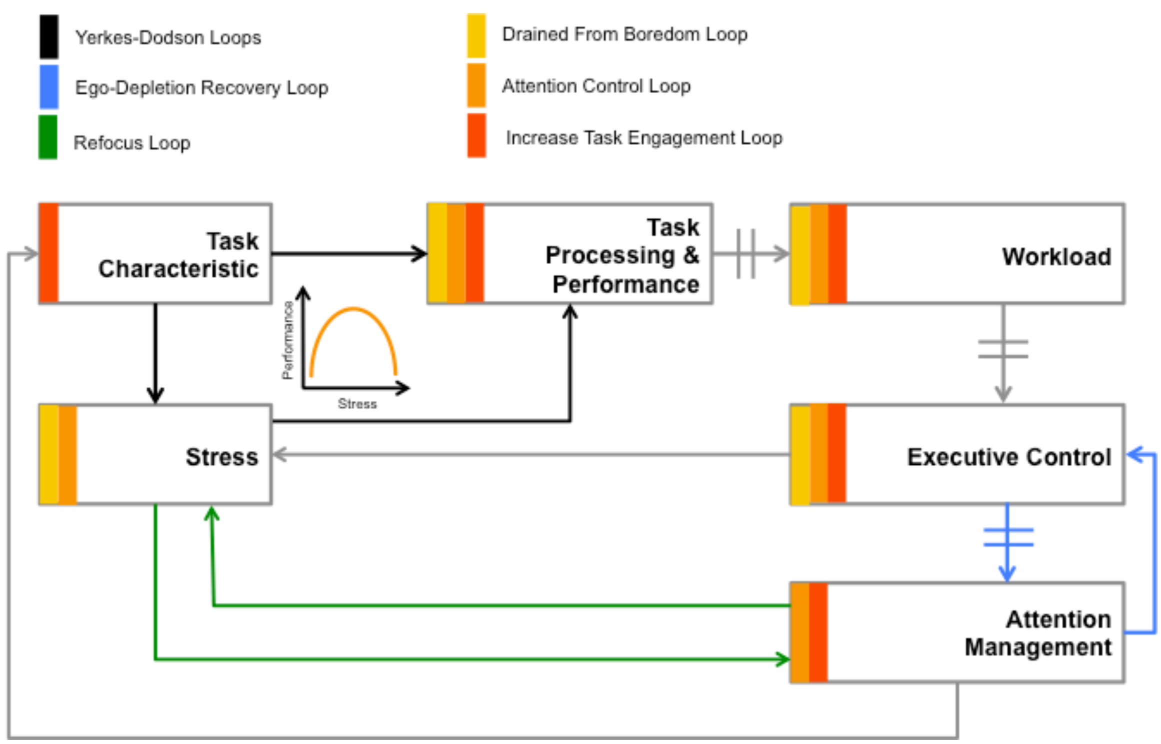


Figure 3-4: Model Overview

The Refocus Loop, as highlighted with green arrows in Figure 3-4, captures an increased stress reaction as task load increases. Intuitively, when the task is more demanding, people devote more attention to it. Theoretically, this corresponds to the alerting and orienting function of attention. When there is a warning cue, the brain transitions from the resting state to an alert state that involves preparation for detecting and responding to a signal. Attention is then oriented to sensory inputs with high priorities (Petersen and Posner 2012; Posner and Petersen 1990).

The Drained from Boredom Loop models the depletion of executive control resource under low task load. Components involved in this loop are highlighted in yellow. When the task load is low, human operators do not have much to do other than monitor the system. This results in low workload, as well as boredom or passive fatigue. When human operators are bored, their will power is drained as part of the executive control resources. When an emergent event happens, the stress level can be even higher with the reduced resources. The change in stress level impacts task processing and performance as discussed previously. If the stress level is too high, performance will decline.

The Attention Control Loop works similarly to the Drained from Boredom Loop with the additional component of Attention Management, as highlighted in orange in Figure 3-4. When the executive control resources are drained, it becomes more difficult to resist distraction temptations. As a result, the attention on the primary task is reduced. This will negatively impact performance.

The last loop is Increase Task Engagement Loop as highlighted in red in Figure 3-4. When the task load is low, some people may try to combat boredom by increasing their task engagement and generating new tasks or interacting with the system even when it is not required. This balances the negative impact of boredom and distraction, and can be good for task performance. However, whether this loop is activated depends on individual strategies in coping with boredom as well as the system design. If the system does not allow additional human interaction under low task load, the ability to increase task engagement is limited.

The model was developed with several iterations. The full diagram of the final model is presented in Figure 3-5. Details of the five interconnecting modules and the equations are described in the following sections.

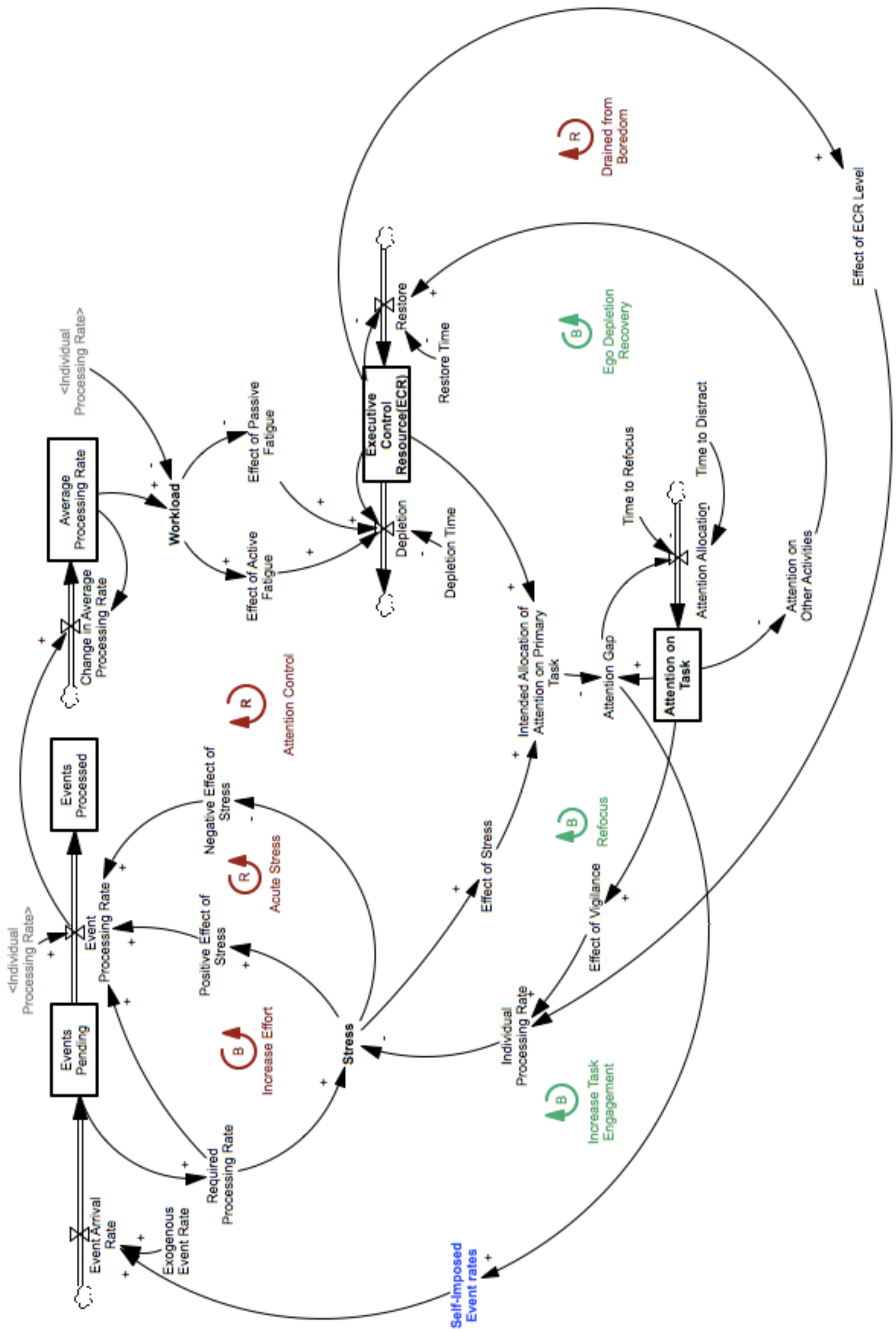


Figure 3-5: Performance and Attention in Low-task-loading Model

3.3.2 Task Characteristics, Processing and Performance

The first module of the SD model is the task characteristics, processing and performance module. This module aims to model task processing process and performance metrics in a generic way. It includes two stocks (Events Pending, Events Processed) and two flows (Event Arrival Rate, Event Processing Rate). Depending on the nature of the task, events can arrive in various patterns and require different level of human effort, as summarized in Table 3-1. Here an event is defined as an occurrence in the system that requires action from a human to process.

Table 3-1: Task Types and Example

Frequency	Mental Effort	Examples
High	High	Air Traffic Control
High	Low	Manufacturing, Long duration driving
Low	High	Rare accidents, Alarms
Low	Low	Simple approval of infrequent actions

In air traffic control tasks, for example, events can happen frequently, particularly at busy airports, and require high human effort level. Under these situations, the novelty and uncertainty of events are both high. People need to response to each event with high mental effort. For repetitive tasks in manufacturing, events arrive continuously with a high frequency, but require low levels of human mental effort to process. Often, these tasks rely more heavily on physical effort than mental effort. In addition, familiarity with the task and experience usually decrease the mental effort required, achieving a state of automaticity (Logan 1985). People in automaticity states perform tasks quickly, effortlessly, and have much unused cognitive capacity. Manufacturing workers and truck drivers often can perform their tasks while having their mind wander.

There are also tasks for which events happen infrequently. Human supervisory control tasks can be categorized as low event frequency and high human effort. While many events can be processed by automation without the need for a human operator, operators are required to monitor the system status and process any events beyond the capability of automation, such as alarms, accidents, and emergencies. Such events do not happen frequently, but often require complex information processing and decision-making. Tasks with low event frequency and low human effort are rare for primary tasks. Examples are renewing one's driver's license once in a few years with no change in health condition, following an infrequently used but simple procedure in organization, etc. For human automation interaction, an automated system that occasionally requires human consent for task execution is in this category.

By setting different parameters for task arrival rates and human mental effort level, tasks described above can be modeled in a simplified way. For the scope of this research, the focus is on low task loading automated environments in which events often happen infrequently. However, the model allows adaptation to other tasks as well.

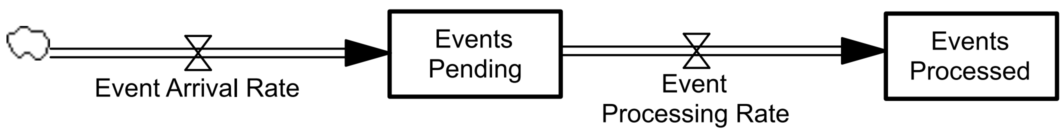


Figure 3-6: Task Characteristics, Processing and Performance

Previous research shows that human supervisory control processes can be modeled as queues (Fei et al. 2014; Mkrtchyan 2011). In a SD model, stocks and flows are used to model events accumulation and processing similar to a queuing model, as shown in Figure 3-6. Event arrival is represented as a flow variable, labeled the *Event Arrival Rate*. It is a combination of exogenous and endogenous event arrival rates. Exogenous events happen when there are external threats, or tasks beyond the capability of automation. Endogenous events are generated by human operators, which are also called Self-Imposed Events in the model. Together,

$$Event\ Arrival\ Rate(t) = Exogenous\ Event\ Rate(t) + Self\ Imposed\ Event\ Rate(t) \tag{1}$$

Events are not processed instantaneously but, instead, accumulate in the stock *Events Pending*. The stock represents the accumulation of events that have arrived but have yet to be processed. Formally,

$$Events\ Pending(t) = \int_t \left[Event\ Arrival\ Rate\ (s) - Event\ Processing\ Rate\ (s)\right]ds + Events\ Pending\ (t_0) \tag{2}$$

Events accumulate because the rate at which they arrive may exceed the rate at which they are processed. The rate at which events are processed is represented by an outflow from *Events Pending*, named *Event Processing Rate*. Event Processing Rate is affected by many factors, including exogenous factors such as task requirement and individual human capability, as well as endogenous factors such as stress level, adequacy of cognitive resource, and attention state. The impacts of these factors are introduced in later sections. Together, these factors determine how fast an event can be processed. For this general model, *Event Processing Rate* or *Net Event Processing Rate* also incorporates the effect of errors and delayed responses. If an error occurs during the event processing and needs to be corrected, this event will remain in the stock until it is discovered and corrected. Delayed response happens when events are not noticed immediately at their arrival due to distraction, lack of vigilance, etc. In either case, *Net Event Processing Rate* is reduced. For a more elaborate model calibrated to a specific task, errors and delayed response can be separated and represented with additional stocks, flows and variables.

Events Processed is the accumulation of events at the completion of processing. This level is often used as a performance measure. The equation is

$$Events\ Processed(t) = \int_t [Event\ Processing\ Rate\ (s)]ds\ +\ Events\ Processed\ (t_0) \tag{3}$$

3.3.3 Stress

The Stress module is connected with the previous module related to task processing. It has five main variables: Required Processing Rate, Individual Processing Rate, Stress, Positive Effect of Stress and Negative Effect of Stress. The amount of accumulated events creates stress, which includes time pressure and perceived performance gaps. Time pressure usually results from the requirement to process the events within a certain time. For goal-directed tasks, the inability of the individual to perform the required actions also introduces stress. In flow theory, anxiety happens when human skills are less than the task challenges (Csikszentmihalyi 2014). If human skill level is higher than the task challenges, people feel calm or bored. Together, a high level of stress arises when the events accumulated need to be processed within a given time and the individual capability is inadequate to do so. The stress level is low when the task is not emergent and its difficulty is within the capability of the individual.

In experiments, stress or arousal level is often an independent variable that remains constant within a given treatment. However, in many real-world situations, stress or arousal level changes overtime. Thus, stress is modeled as the ratio between *Required Processing Rate* and *Individual Processing Rate*. Equation (4) reflects time pressure, and Equation (5) reflects the impact of performance gap. When *Required Processing Rate* equals *Individual Processing Rate*, *Stress* equals a baseline value of one. When *Required Processing Rate* is higher than *Individual Processing Rate*, it means the required task load is higher than the individual capability. Hence, *Stress* takes a value larger than one, indicating a high stress level. Similarly, when *Required Processing Rate* is lower than *Individual Processing Rate*, *Stress* takes a value smaller than one, indicating a low stress level.

$$Required\ Processing\ Rate(t) = Events\ Pending(t)/Required\ Processing\ Time \tag{4}$$

$$Stress(t) = Required\ Processing\ Rate(t)/Individual\ Processing\ Rate(t) \tag{5}$$

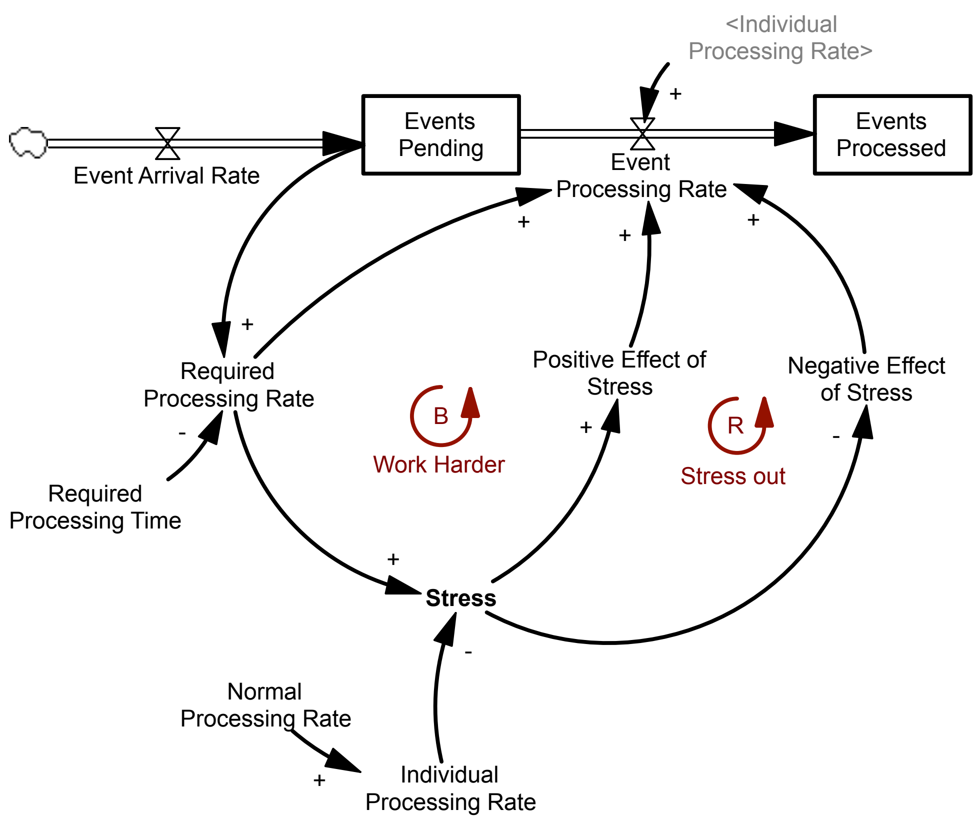


Figure 3-7: Stress and the Impact on Performance

Stress level influences task processing and task performance. Yerkes and Dodson (1908) found that performance in various tasks depended on the level of arousal or stress imposed. Both low and high levels of arousal yield low performance. Best performance is achieved when the person is neither under or over loaded. This inverted U-shaped relationship between stress and performance has been replicated in physical and cognitive tasks beyond the original experiments (Fisher 1986). Despite the ambiguity or disagreement over what constitutes arousal and why such a relationship exists, the Yerkes-Dodson Law shows robustness in various settings (Teigen 1994).

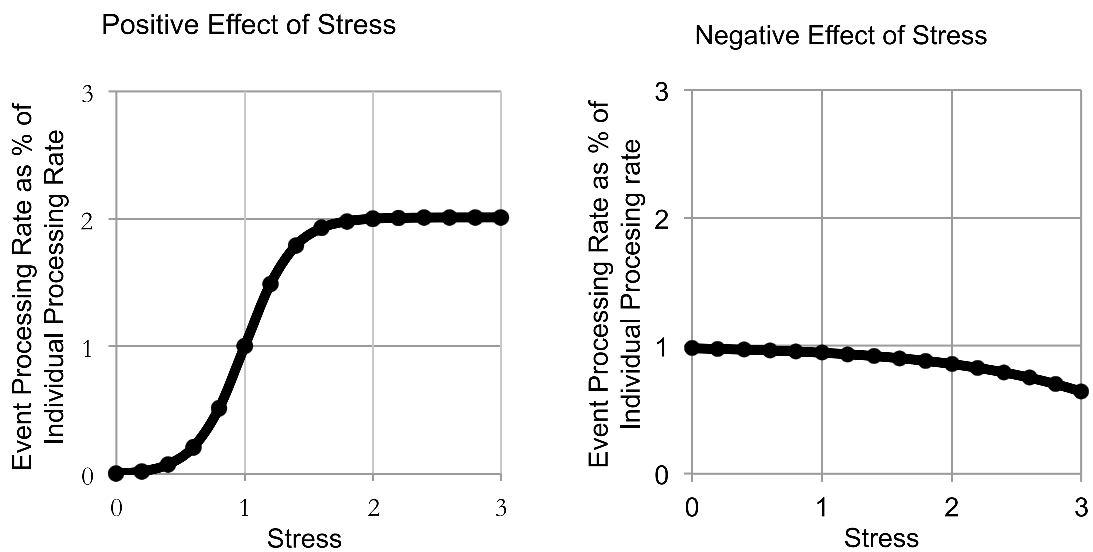
The model captures the impact of stress on performance with two feedback loops, as shown in Figure 3-7. The positive and negative effect of stress are separated into two loops for two reasons: a) to ensure that all causal links have unambiguous polarity, b) a U-shaped relationship indicates the presence of multiple causal pathways between the input and output (Sterman 2000). In SD models, it is a common practice to separate U-shaped relationship into separate loops with unique, unambiguous polarity (Sterman 2000). The *Positive Effect of Stress* is conceptualized as a consequence of effort regulation. Humans tend to spend as little effort as possible in performing mental tasks (Kahneman 2011). Increased stress signals the necessity to increase effort level. In order to maintain the desired level of performance, people can often adapt dynamically to changing demands through effort regulation (Hancock and Warm 2003). Driving research has shown, for example, that people's performance decreases in under load conditions due to a loss of task-directed effort (Matthews and Desmond 2002). Motivated by these findings, the balancing loop of *Increase Effort* in Figure 3-7 captures the increase of performance when the stress level rises. An S-shaped function was used in a previous SD model that incorporate the Yerkes-Dodson law to capture the impact of stress on human decision-making regarding accidents and disasters (Rudolph and Repenning 2002). Taking a similar approach, Equation (6) models the positive effect of stress. This equation was derived based on two key points on the S-curve: 1) when *Stress* = 1, *Positive Effect of Stress* = 1; 2) when *Stress* = 0, *Positive Effect of Stress* = 0. The relationship is shown graphically in Figure 3-8(a).

$$Positive\ Effect\ of\ Stress\ (t) = \frac{2(1+e^{k_1})}{(e^{k_1}-1)[1+e^{-k_1(Stress(t)-1)}]} - \frac{2}{e^{k_1}-1} \tag{6}$$

$$Negative\ Effect\ of\ Stress\ (t) = \begin{cases} 1 & \text{for } Stress(t) \leq 1 \\ 1 - \dfrac{c_2}{1 + e^{-k_2(Stress(t)-2)}} & \text{for } Stress(t) > 1 \end{cases} \quad (7)$$

Where $k_1 \geq 1, k_2 \geq 1, c_2 > 0$

$$Event\ Processing\ Rate(t) = \quad (8)$$
$$MIN(Required\ Processing\ Rate(t), Individual\ Processing\ Rate(t) *$$
$$Positive\ Effect\ of\ Stress(t) * Negative\ Effect\ of\ Stress(t))$$



(a) (b)

Figure 3-8: Positive and Negative Effects of Stress on Event Processing Rate

The *Negative Effect of Stress* reflects the limitation of human cognitive capability. Human mental energy, memory and attention are all limited resources. When the task requirement exceeds cognitive capability, performance decreases because the available resources are inadequate to cope with the situation. Individuals may simply ignore some tasks, decrease their effort on low priority tasks, or reduce the quality of work. The reinforcing loop of *Acute Stress* captures the decrease of performance when the stress level is too high. The negative effect of stress is modeled using Equation (7). The relationship is shown more clearly in Figure 3-8(b). *Stress*, *Positive Effect of Stress*, and *Negative Effect of Stress* are all dimensionless variables. Taken together, *Event Processing Rate* equals *Individual Processing Rate* when the stress level equal to one, which is the baseline stress level. When the stress level is less than one, events are processed at a rate less than the *Individual Processing Rate*. When the stress level is more than one, *Event Processing Rate* increases first and then declines.

3.3.4 Workload

This module in the SD model captures the workload of an operator as affected by task demand and individual capability. It has four variables: *Average Processing Rate*, *Change in Average Processing Rate*, *Time Window*, and *Workload*.

Workload is a mental construct that reflects the task demand coupled with the capability of the operator to respond to the task demand (Cain 2007; Moray 2013). Task demand is affected by factors including task design, task procedure, environment and situational factors. Although these cannot all be captured in the model, we try to capture the major component of task demand using an average of task processing rate. Task processing rate not only reflects whether the operator is idle or busy, but also how fast the task needs to be processed. It is an objective measure of task load. When there is no task, the task processing rate equals zero, resulting in zero workload. A higher task processing rate indicates a higher workload, and vice versa.

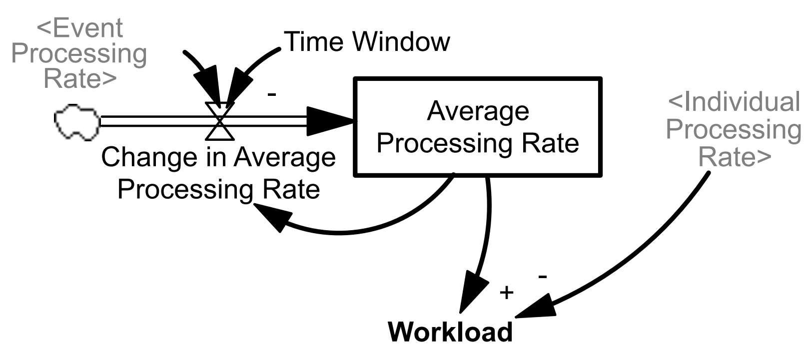


Figure 3-9: Perceived Workload

Exponential smoothing was used to calculate the *Average Processing Rate*, as shown in Figure 3-9. People don't update their perception instantly. Instead, there is often a delay in updating when a gap in perception is identified. In addition, exponential smoothing incorporates recent changes in the target variable better than a moving average because recent changes are given larger weights. The equation for *Average Processing Rate* is as below:

$$Average\ Processing\ Rate(t) = \int_t Change\ in\ Average\ Processing\ Rate(s)ds + Average\ Processing\ Rate(t_0) \tag{9}$$

$$Change\ in\ Average\ Processing\ Rate(t) = (Event\ Processing\ Rate(t) - Average\ Processing\ Rate(t))/(Time\ Window) \tag{10}$$

$$Workload(t) = \frac{Average\ Processing\ Rate(t)}{Individual\ Processing\ Rate(t)} \tag{11}$$

Workload as a subjective measure is affected by both task load and individual capability. In the SD model, *Workload* is then defined as the ratio between *Average Processing Rate* and *Individual Processing Rate*, as in Equation (11). This reflects inadequacy or surplus of individual capability regarding the recent task-processing load. There are other ways to model workload, such as utilization level in terms of busy time over the sampling time period (Clare 2013). However, utilization level does not reflect the level of cognitive effort people devoted. In a high workload task scenario, ignoring effort level may not have a large impact, because people are likely to devote their full effort most of the time. When workload includes primarily monitoring tasks, which can be high in mental workload, incorporating effort level into workload is critical. Workload is lower when working at a comfortable pace comparing to hustling, and when the individual is more capable of handling the tasks.

In summary, task processing affects the workload of individuals. Given the same task, an individual may experience different levels of workload depending on his/her individual capability. Studies have shown that human cognitive workload affects both human and system performance (Clare 2013; Cummings et al. 2013; Wierwille et al. 1985). In the PAL model, the influence of low workload is captured by the Drained from Boredom Loop. Workload affects the level of fatigue and consumption of cognitive resources, which then impacts stress level, and ultimately task processing per the Yerkes-Dodson Law relationship (Figure 3-4).

3.3.5 Executive Control

This module captures the influence of workload on executive control resource. It includes the following variables: Effect of Active Fatigue, Effect of Passive Fatigue, Executive Control Resource, Maximum Level, Minimum Level, Depletion, Recovery, Depletion Time, Recovery Time, and Personal Precursors.

Humans have limited cognitive resources. These resources may be defined as reservoirs of "fuel" or "energy" for cognitive processes. While there are many types of cognitive resources, executive control resource is critical. Executive control resource is allocated to a variety of specific processes involved in the control of cognition and goal-directed behavior, such as information processing, attention allocation, self-regulation, decision-making, etc. (Gross and Grossman 2010). Shifting attention to salient items or events, sustaining attention despite distraction or interference, and overriding automatic responses with more appropriate behaviors are all examples of executive functions. *Executive Control Resource (ECR)* is modeled as a stock ranging from zero to one to reflect its limited capacity, as shown in Figure 3-10. *Executive Control Resource* is initialized at its maximum level, which is determined by individual differences such as boredom proneness, sleep quality, gaming experience, etc. A series of laboratory studies suggests that self-control depends on a limited executive control resource, which is depleted with acts of self-control and restored with rest and positive affect (Baumeister 2002). In the model, outflow *Depletion* drains *Executive Control Resource*, and inflow *Recovery* increases it. Formally,

$$Executive\ Control\ Resource(t) = Executive\ Control\ Resource(t_0) - \int_t Depletion(s)ds + \int_t Recovery(s)ds \qquad (12)$$

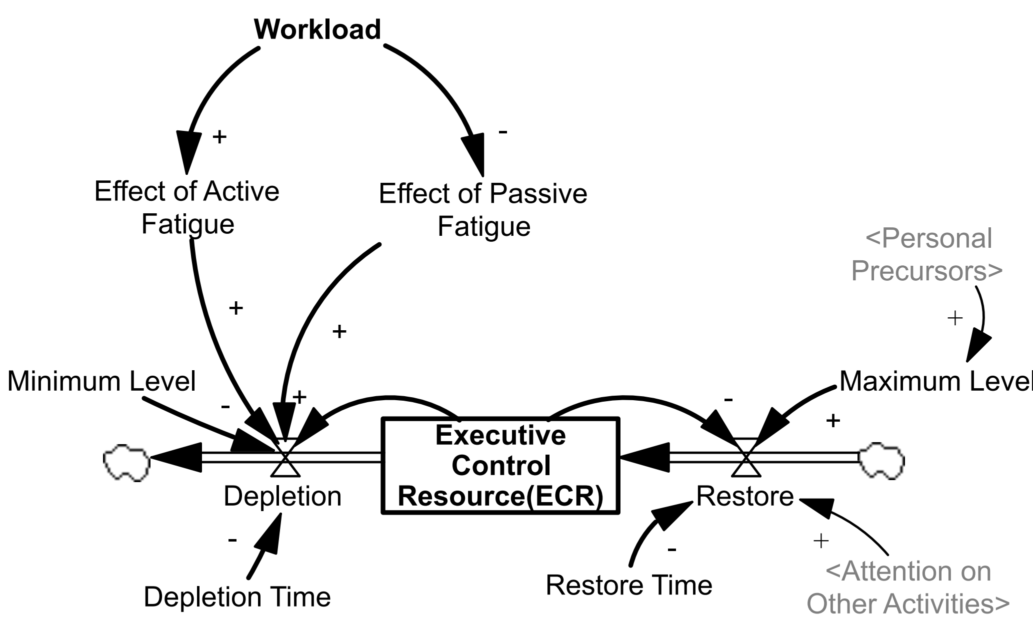


Figure 3-10: Executive Control Resources

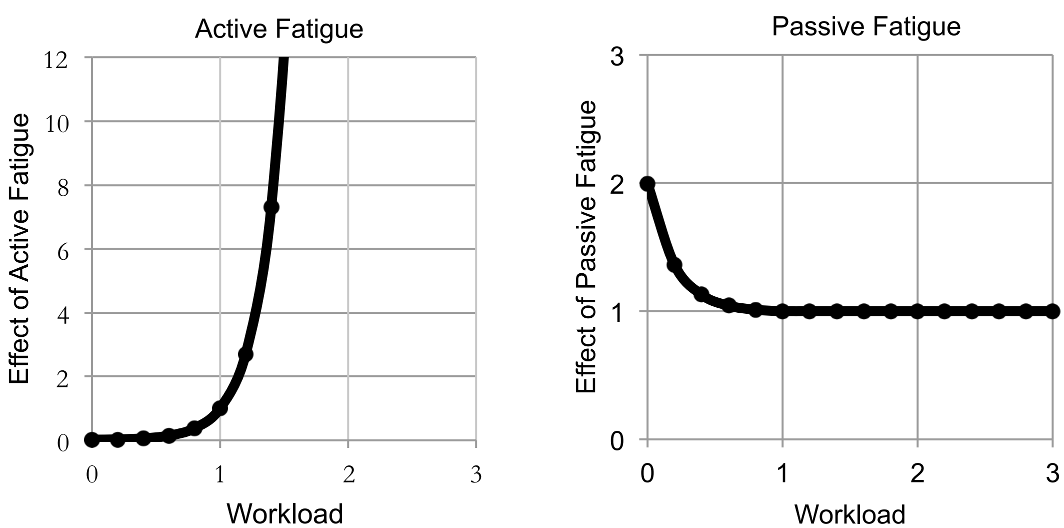
Executive control resource depletion, or self-control exertion, could lead to self-control failure at a later time (Inzlicht et al. 2014). Such depletion could happen because of fatigue and prior regulatory demands (Muraven et al. 1998). Fatigue here refers to *Effect of Active Fatigue* in the model resulting from high workload. Prior regulatory demands correspond to *Effect of Passive Fatigue* in the model. In the SD model, *Depletion* is influenced by three processes. In addition to active and passive fatigue, depletion is also limited by the current executive control resource level. As the level of executive control resource decreases, depletion slows down because there is less to drain. In the extreme case, *Depletion* equals zero when *Executive Control Resource* is zero.

$$Effect\ of\ Active\ Fatigue\ (t) = \frac{L_1}{1+e^{-k_5(Workload(t)-c)}} + L_2 \quad (13)$$

where $L_1 = \dfrac{(1-m_5)(1+e^{k_5c_5})(1+e^{-k_5+k_5c_5})}{e^{k_5c_5}(1-e^{-k_5})}$

$$L_2 = m_5 - \frac{L_1}{1+e^{k_5c_5}}$$

$$Effect\ of\ Passive\ Fatigue\ (t) = MAX(1, 1 + c_6e^{-k_6Workload(t)} - c_6e^{-k_6}) \quad (14)$$



(a) (b)

Figure 3-11: Effect of Active Fatigue and Passive Fatigue

High workload introduces *Active Fatigue*, which results in faster resource depletion. Active fatigue is derived from continuous and prolonged, task-related perceptual-motor adjustment

(Hancock and Desmond 2001). In studies involving such type of tasks, it was found that fatigue increases over time during the task, referred to as time-on-task effect (Coull et al. 1998; Lim et al. 2010). While active fatigue is related to factors such as sleep deprivation, nutrition, and depression, workload is a significant predictor for *Active Fatigue* (MacDonald 2003). Higher workload results in higher level of fatigue. Active fatigue could result in diminished capacity for work, disinclination to apply effort to the task, and perceived reductions in personal efficiency difficulties in concentrating and focusing attention (Matthews and Desmond 2002). The effect of *Active Fatigue* is modeled using a S-shape function as shown in Figure 3-11(a) and Equation (14). Following this equation, the *Effect of Active Fatigue* equals one when *Workload* is one. When *Workload* is zero, the *Effect of Active Fatigue* equals a small positive number.

While active fatigue often happens under high workload, low workload could impose regulatory demands. In boring work environments, people need to resist the temptation of mental leisure activities such as mind wandering and distraction. This consumes executive control resources, making it even harder to resist the temptation in later time. The third process that depletes the executive control resource is *Passive Fatigue*. *Passive Fatigue* develops when the task requires system monitoring with either rare or even no overt perceptual-motor response requirements (Hancock and Desmond 2001). The depletion is faster in low workload scenarios compared to those of optimal workload due to frustration, boredom, ego depletion and self control failure (Baumeister 2002). In a study with Navy patrol vessel crewmembers, fatigue was found to be associated with low workload vigilance tasks at the beginning of the patrol (passive fatigue) but with high workload by the end of the patrol (active fatigue), following a U-shaped curve (Grech et al. 2009). Researchers have found that brain is more active at rest than it is in a range of explicit tasks based on brain imaging results (Morcom and Fletcher 2007). The effect of *Passive Fatigue* is then modeled using the function shown in Figure 3-11(b) and Equation (14). The effect of *Passive Fatigue* equals one when the *Workload* is one or higher. *Passive Fatigue* increases as *Perceived Workload* reduces to less than one. Taken together, the equation for *Depletion* is presented in Equation (15). *Effect of Active Fatigue* is multiplied with *Effect of Passive Fatigue* to formulate the U-shaped curve. Under high workload, the overall fatigue equals active fatigue. Under low workload, passive fatigue is dominating.

$$Depletion(t) = \frac{ECR(t) - Minimum\ Level}{Depletion\ Time} * Effect\ of\ Active\ Fatigue(t) * Effect\ of\ Passive\ Fatigue(t) \quad (15)$$

The *Restore* process happens when attention is devoted to other activities such as task-unrelated thought and secondary tasks. For example, when engaged in mind wandering, an individual may not feel bored at the time (Eastwood et al. 2012). Research shows that secondary task engagement reduces boredom for drivers during long drives (Atchley and Chan 2011; Oron-Gilad et al. 2008). Any new stimulus that is different from the primary task could mitigate the negative affect of boredom and frustration, resulting in a recovery of executive control resource.

$$Restore(t) = \frac{Maximum\ Level - ECR(t)}{Restore\ Time} * Attention\ on\ Other\ Activities(t) \quad (16)$$

Individuals differ in their tolerance for boredom. Personal precursors such as boredom proneness and sleep quality influences how easily one gets bored and how much self-regulatory power one has to stay focused when bored. Boredom proneness is found to relate to an individual's ability to manage sustained attention tasks (Farmer and Sundberg 1986), impatient behavior, distraction, sensation seeking, impulsiveness, and work performance (Dahlen et al. 2005; Kass and Vodanovich 1990; Vodanovich et al. 1991). Game-playing also correlates with boredom (Zhou 2010). We capture these individual differences in the variable *Personal Precursors*. *Maximum Level* of *ECR* is influenced by *Personal Precursors*. The equation for *Maximum Level* and *Personal Precursors* (Figure 3-10) will be discussed in models for specific task environment in Chapter 6.

As many cognitive processes use executive control resource, the level of ECR could influence performance directly. Research has shown that fatigue could result in sub-optimal performance in terms of increased errors, delayed response times, and longer task processing times (Lorist et al. 2000; van der Linden et al. 2003). Vigilance performance has been shown to be worse under increasing fatigue (Denisco et al. 1987). In the PAL model, fatigue leads to decreased level of ECR, which then influences performance. The impact of decreased level of *ECR* on performance was modeled using a S-Shaped function, with higher resource levels resulting in higher *Individual Processing Rate* (Figure 3-12). The *Effect of ECR* takes a value between zero and one, and is calculated based on Equation (17). Connecting *Individual Processing Rate* with *Stress* and *Performance* closes the loop of Drained from Boredom. This feedback loop represents that a certain level of workload can be beneficial in slowing down resource depletion and reducing stress, while a low level of workload can drain executive control resource and hurt performance.

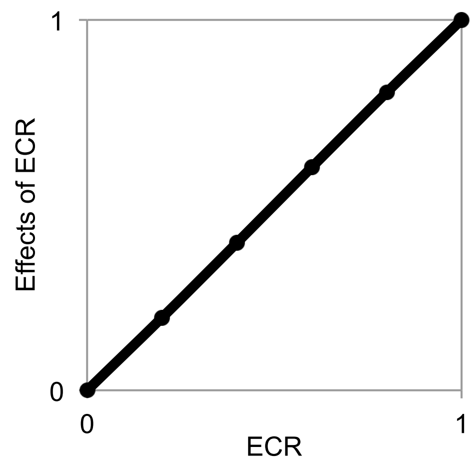


Figure 3-12: Effect of Executive Control Resource (ECR)

$$Effect\ of\ ECR(t) = \frac{(1+e^{-k_4})(1+e^{k_4})}{(e^{k_4}-e^{-k_4})(1+e^{-2k_4ECR(t)+k_4})} - \frac{(1+e^{-k_4})}{(e^{k_4}-e^{-k_4})} \tag{17}$$

3.3.6 Attention Management

The last module captures the attention change of the operator as affected by stress and executive control resource level. It includes the following variables: Intended Allocation of Attention on Primary Task, Attention Gap, Attention on Task, Attention on Other Activities, Attention Allocation, Time to Refocus, Time to Distract, and Effect of Stress.

Executive control resource influences the attention allocation process. Low workload introduces passive fatigue and boredom, which causes difficulties in focusing attention on the primary task. This results in mind wandering or distraction, causing a decrease in vigilance and task performance. The overview of these causal links is shown in Figure 3-13.

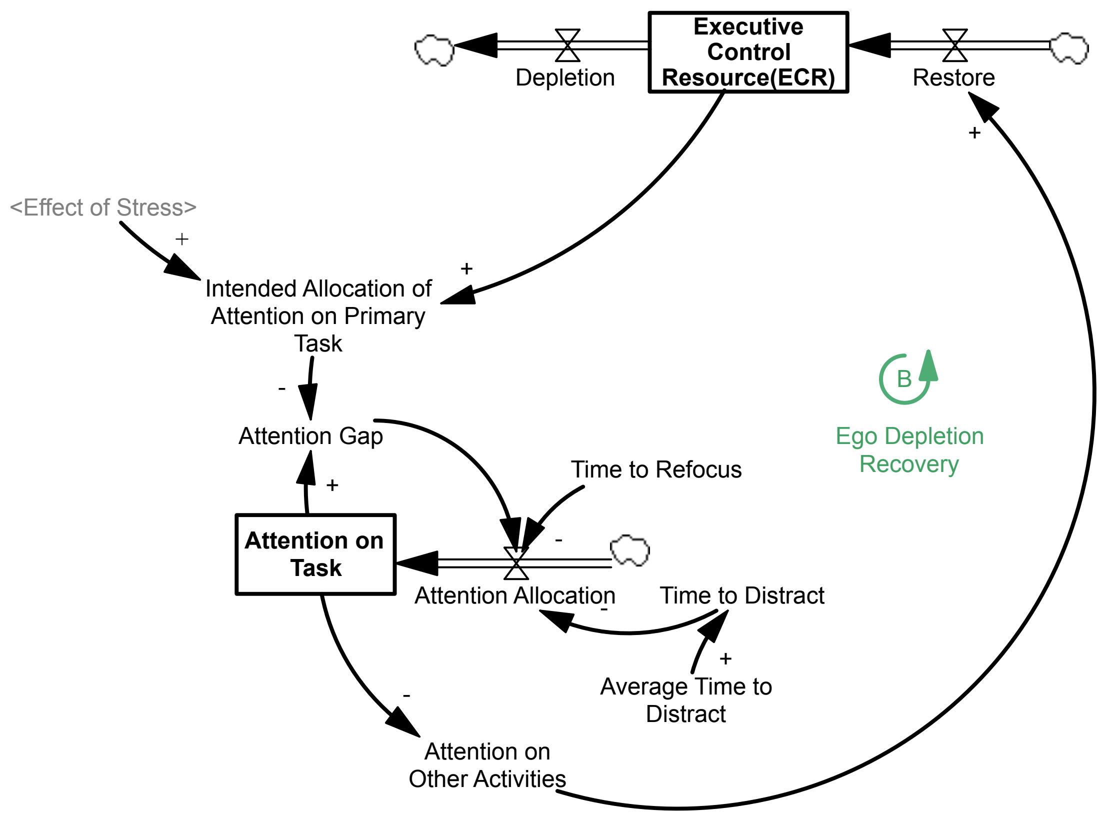


Figure 3-13: Attention Management

The central construct of this process is attention. Sustained attention is the ability to maintain a consistent behavioral response during continuous and repetitive activity (Salvendy 2012). Since attention is a type of cognitive resource with limited capacity, it is modeled as a stock called *Attention on Task*. Distracted Attention is the attention of an individual that shifts from the chosen object of attention onto the source of distraction. When the operator engages in task unrelated thought (TUT) or gets distracted, sustained attention decreases and distracted attention increases. Previous research show that TUTs draws attentional resources away from the primary task, resulting in degraded performance (McVay and Kane 2010). Divided attention refers to the ability to respond simultaneously to multiple tasks or multiple task demands. When the operator engages in secondary tasks that are related to the primary task, sustained attention decreases and divided attention increases. Although distracted attention and divided attention are different cognitive concepts, we combine them together as *Attention on Other Activities* to simplify the model, as the major concern is whether the attention is on the primary task or not. *Attention on Task* and *Attention on Other Activities* both range from zero to one. The sum of these two variables always equals to one.

$$Attention\ on\ Other\ Activities(t) = 1 - Attention\ on\ Task(t) \tag{18}$$

Intended Allocation of Attention on Primary Task is the level of attention that an individual plans to allocate to the task. *Attention on Task* is the actual level of attention being devoted to the task at a specific time. *Intended Allocation of Attention on Primary Task* is affected by two variables as formulated in Equation (19): *Executive Control Resource* and *Effect of Stress*. *Executive Control Resource* affects how much a person can stay focused on the primary task. When people get bored, they may try to increase task engagement by directing their attention towards the primary task intentionally and trying to resist the temptation to get distracted. However, when ECR is depleted, individuals will not be able to sustain their attention (Langner and Eickhoff 2013). Research also shows that mentally fatigued individuals cannot inhibit automatic shifting of attention to irrelevant stimuli (Boksem et al. 2005). This means the *Intended Allocation of Attention on Primary Task* is lowered when ECR is decreased. The second variable that affects *Intended Allocation of Attention on Primary Task* is *Effect of Stress,* which reflects the impact of task demand on attention management, as modeled in Equation (20) and Figure 3-14. Research shows that TUTs increase when there are more unused cognitive resources, and decrease as task load or difficulty increases (McVay and Kane 2010). This means *Intended Allocation of Attention on Primary Task* is increased under high stress level as the result of high task demand.

$$\begin{aligned} &Indicated\ Demand\ for\ Attention\ on\ Primary\ Task\ (t) \\ &\qquad = Min((1 + Effect\ of\ Stress(t)) * Executive\ Control\ Resource(t), 1) \end{aligned} \tag{19}$$

$$Effect\ of\ Stress(t) = \frac{c_7}{1 + e^{-k_7 * Stress(t)}} - \frac{c_7}{2} \tag{20}$$

When the actual attention level does not equal the intended allocation of attention, adjustment in attention allocation happens. Such adjustment takes time. *Attention on Task* is adjusted to *Intended Allocation of Attention on Primary Task* as an information delay process.

Depending on whether attention is shifted away from or back to the primary task, a different delay time (*Time to Distraction* or *Time to Refocus*) is used. The longer a person stays in an under-stimulating environment, usually the harder it is to concentrate. Neuroscience research based on EEG signals shows that the overall level of physiological vigilance decreases over time (Campagne et al. 2004). Research shows that concentration on a task decreases after 10 minutes or even earlier (Szalma et al. 2004). Other research found that people's vigilance begins to decline after 20-30 minutes for a task that requires sustained attention (Wickens et al. 2011). This is captured by the variable *Time to Distract* in Figure 3-13.

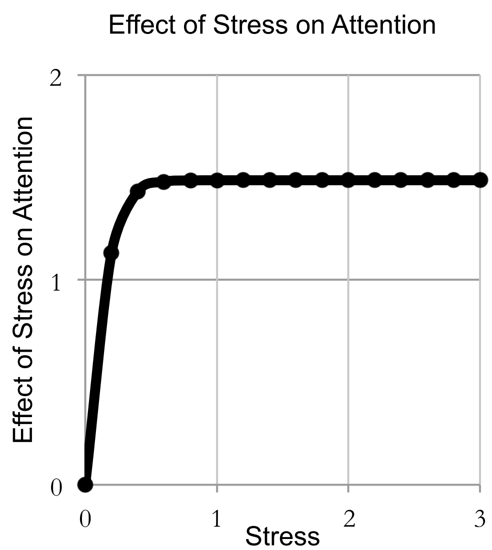


Figure 3-14: Effect of Stress on Attention

Attention can be allocated back to the primary task when task demand increases. However, one cannot shift their attention immediately as a new task arrives. Usually, people need to spend extra time to reallocate their attention and regain situation awareness. Research shows that task switching is costly as reflected in the substantially slower responses and, usually, more error-prone immediately after a task switch (Monsell 2003). The time required to refocus is modeled as *Time to Refocus*, which is usually shorter than the *Time to Distract*.

$$Attention\ Gap(t) = Intended\ Allocation\ of\ Attention\ on\ Primary\ Task(t) - Attention\ on\ Task(t) \tag{21}$$

$$Attention\ Allocation(t) = \frac{Attention\ Gap(t)}{Time\ to\ Refocus(or\ Time\ to\ Distract)} \tag{22}$$

$$Attention\ on\ Task(t) = Attention\ on\ Task(t_0) + \int_t Attention\ Allocation(s)ds \tag{23}$$

$$Individual\ Processing\ Rate(t) = Normal\ Processing\ Rate(t) * Effect\ of\ ECR(t) * Effect\ of\ Vigilance(t) \tag{24}$$

$$Effect\ of\ Vigilance(t) = \frac{(1+e^{-k_3})(1+e^{k_3})}{(e^{k_3}-e^{-k_3})(1+e^{-2k_3 ECR(t)+k_3})} - \frac{(1+e^{-k_3})}{(e^{k_3}-e^{-k_3})} \tag{25}$$

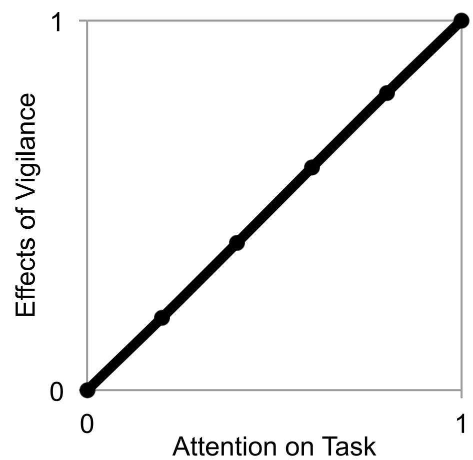


Figure 3-15: Effect of Vigilance

The final piece of the PAL model connects attention with vigilance and task performance. Vigilance describes the state of readiness to detect and respond to stimulus changes that are barely detectible, or which occur infrequently or at irregular intervals (Ballard 1996). Sustained attention is a necessary condition for vigilance (Berka et al. 2007). It is reasonable to assume that higher levels of attention correspond to higher levels of vigilance. Vigilance affects task performance since a state of high alertness enables faster responses (Posner and Petersen 1990). In another study, task engagement, which includes energetic arousal, motivation and concentration, was found to correlate with vigilance task performance (Matthews et al. 1999). We modeled *Effect of Vigilance* with a S-Shaped function that increases as *Attention on Task* increases, as in Equation (25). The curve is linear when k_3 equals one. A decreased vigilance level has a negative impact on performance. The vigilance decrement is commonly measured in terms of missed signals, longer reaction times, and generally poorer performance than can reasonably be expected (Davies and Parasuraman 1982; Warm Matthews et al. 2008). We model this by multiplying *Individual Processing Rate* with the level of vigilance.

When there is excessive attention allocated to the task, people may increase task engagement. For example, human operators may interact with the system beyond the basic task requirement. This is modeled in Equation (26).

$$Self\ Imposed\ Event\ Rate(t) = Baseline\ Self\ Imposed\ Event\ Rate * MAX(-Attention\ Gap(t), 0) \tag{26}$$

3.3.7 Summary

The PAL Model integrates multiple constructs relating to cognitive processes under low task load using five modules: Task Characteristics, Processing and Performance, Stress, Workload, Executive Control, as well as Attention Management. Rather than treating them as separate factors, the interactions among these modules are modeled through six feedback loops: Yerkes-Dodson Loops, Ego-Depletion Recovery Loop, Refocus Loop, Drained from Boredom Loop, Attention Control Loop, and Increase Task Engagement Loop.

The model provides two key output variables that a system designer might be interested in: attention and human performance. Different from a static model, dynamic changes of attention and performance can be examined. A designer could investigate human performance when working with different levels of automation by varying the required interaction frequency. In addition, the effectiveness of using designs and policies to improve attention management and performance can also be evaluated using the model. Using this model to aid the design process could reduce the need for costly and time-consuming human-in-the-loop experiments. It also allows the exploration of wider design choices than is possible through prototyping or experimentation. Before the model is put into use, it needs to be tested and validated. Model Structure Tests are introduced in Section 3.4. Replication and prediction tests are introduced in Chapters 4 to 6.

3.4 Model Structure Tests

The main steps of system dynamics modeling includes problem articulation, formulation of dynamic hypothesis, formulation of a simulation model, testing, policy design and evaluation

(Sterman 2000). In this section, the model is tested using the following methods: Boundary Adequacy Test, Dimensional Consistency Test, Structure Assessment Tests, and Extreme Condition Tests (Sterman 2000). Although only the final results of these tests are presented, these tests are used throughout the modeling processes for several iterations to correct errors and improve the model.

3.4.1 Boundary Adequacy Test

The boundary adequacy test asks whether the model is appropriate for the intended purpose and whether the model includes all the necessary variables and structures (Sterman 2000). Boundary adequacy is inspected by using a model boundary chart as shown in Table 3-2, review of relevant literature, and discussion with experts.

The purpose of this research is to capture attention management and its impact on performance in low task load automated environments using a generalized model. As reviewed in Chapter 2, the key constructs related to this are task characteristics, attention lapse, fatigue, boredom, frustration and complacency, attention management, as well as performance impact and perceived workload. All of these are captured in the SD model. While the representation of these constructs is straightforward, boredom is modeled in the *Effect of Passive Fatigue*. Passive fatigue is the depletion of cognitive resources due to monotony and boredom (Desmond and Hancock 2001). The emotional aspect of boredom, frustration and complacency is also captured in the change of *Executive Control Resource*. These factors are captured endogenously as they are influenced by other variables in the model.

The characteristics of the task, system, and human operators are modeled as exogenous variables as they are not influenced by other variables in the model. In addition, they remain relatively static over the time horizon of the model.

In order to build a generalized model, some details need to be excluded, such as details of tasks, user interface, etc. In addition, human attention is a very complex process, as many cues can affect attention. Since this model's focus is on the change of attention over a long duration, details of attention states such as the difference between distraction and TUTs, and difference between visual and auditory attention are excluded. Keeping a balance between simplicity and thoroughness of the model is important. While the current model suits our purpose, it can be expanded to include more details.

Table 3-2: Model Boundary Chart

Endogenous	Exogenous	Excluded
Task Processing/Performance: • Events Processing Rate • Events Processed • Individual Processing Rate **Stress** **Workload** **Executive Control:** • Executive Control Resource • Depletion • Restore **Attention:** • Attention on Task • Attention on Other Activities • Attention Allocation	**Task Characteristics:** • Event Arrival Rate • Normal Processing Rate • Required Processing Time **Personal Precursors:** • Boredom Proneness • Sleep Quality **System Design/Policy:** • Sources of Distraction • Attention Alert **Human Time Constants/Delays:** • Time to Distract • Time to Refocus • Depletion Time • Restore Time **Initial Conditions:** • Initial Attention on Primary Task • Initial ECR **Non-linear Relationships:** • Effect of Active Fatigue • Effect of Passive Fatigue • Positive Effect of Stress • Negative Effect of Stress • Effect of ECR Level • Effect of Vigilance • Effect of Stress on Attention	**Details of Tasks:** • Details of Unexpected Events • Details of Automated Assets **Details of Control Interface and Task Processing:** • Operation of Automation • Following Safety Procedure • Communication with Control Center (Chat Messages) **Details of Attention State:** • Differentiation between Distraction and TUTs • Distraction Sources • Differentiation between visual and auditory attention **Cascading Errors** **Environmental Effects**

3.4.2 Dimensional Consistency Test

Dimensional consistency testing evaluates the units used for each variable and ensures that the units match on each side of every equation. Units errors reveal important flaws in the model structure (Sterman 2000). All of the equations in the SD model passed the dimensional consistency test, both through inspection of all model equations and through the Vensim® unit check function.

3.4.3 Structure Assessment Tests

The purpose of structure assessment tests is to examine whether the model is consistent with knowledge of the real system relevant to the model purpose. For the PAL Model, this means whether model behavior represents realistic human behavior. To investigate this, the diagram and the model equations are carefully inspected by experts. Partial model tests are used to examine the model behavior. This is done in several steps for this model. First, the Yerkes-Dodson Loops were isolated and inspected. The Drained from Boredom and Attention Control Loops were then added individually. Finally, the full model was inspected.

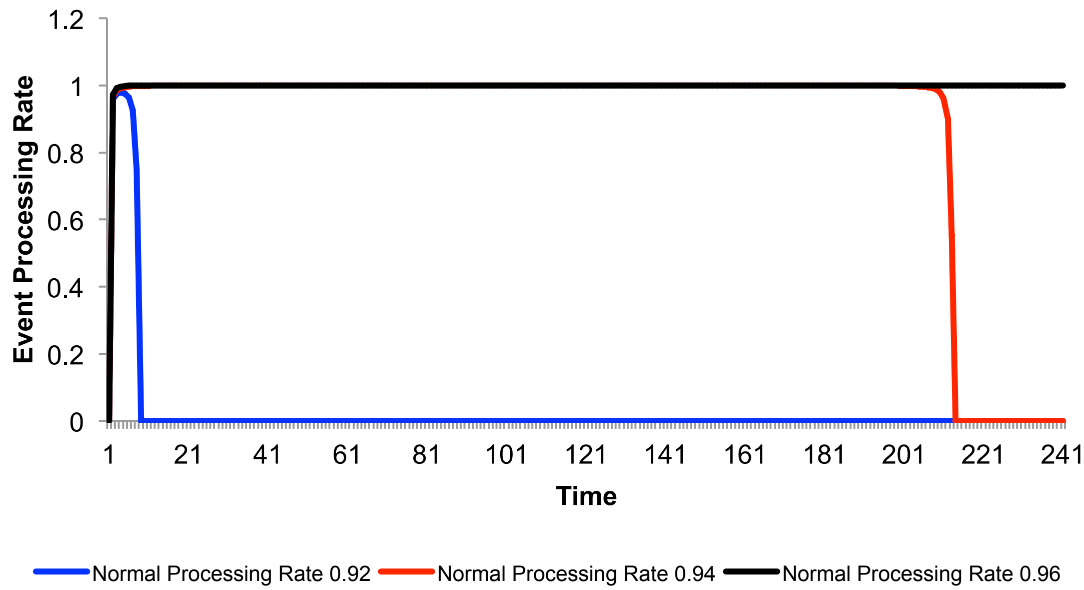


Figure 3-16: Impact of Changes on Normal Processing Rate

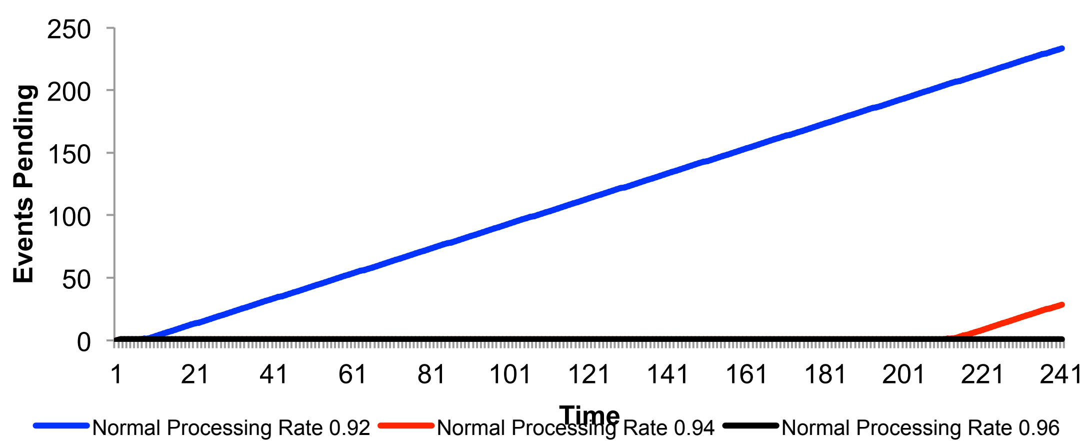


Figure 3-17: Impact of Changes on Events Pending

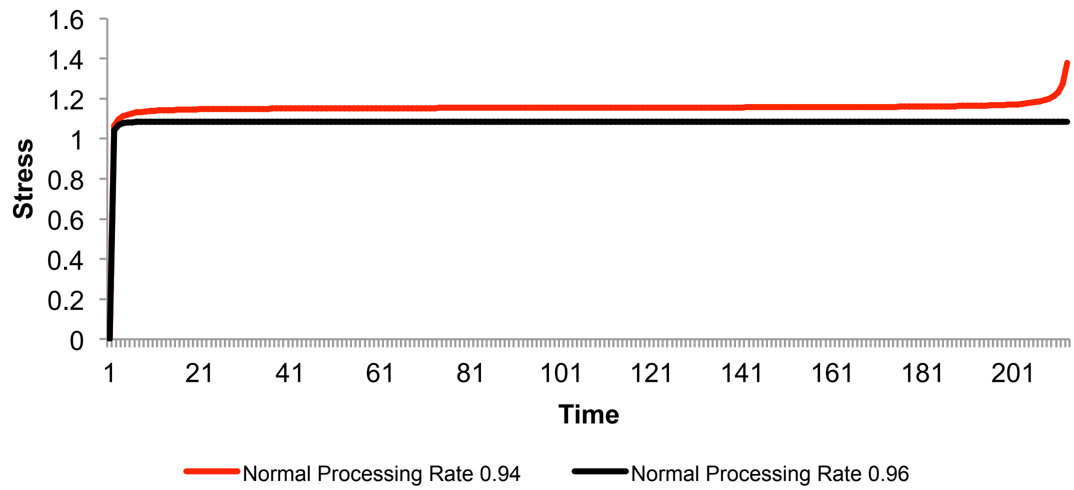


Figure 3-18: Impact of Changes on Stress

In order to test the Yerkes-Dodson loops, the *Event Arrival Rate* was kept as a constant. *Normal Processing Rate* was varied to generate different levels of stress. When *Normal Processing Rate* equaled to *Event Arrival Rate*, the human operator could process all the tasks but would be busy all the time. When *Normal Processing Rate* was larger than *Event Arrival Rate*, tasks would be processed as soon as they arrive. However, if *Normal Processing Rate* was smaller than *Event Arrival Rate*, there may be backlog due to the slow processing. A backlog could then increase the stress level, which influenced the performance following the Yerkes-Dodson law. For simplicity, *Event Arrival Rate* was set to 1 task/minute. *Required Processing Rate* was set to 1 minute. *Normal Processing Rate* was varied between 0.9 task/minute to 1 task/minute. The results are shown in Figure 3-16. When *Normal Processing Rate* equaled to 0.92, *Event Processing Rate* quickly dropped to zero due to the excessive stress from backlog. When *Normal Processing Rate* equaled to 0.94, the drop happened much later because the accumulation of the backlog of tasks was slower, as shown in Figure 3-17. These showed the negative effect of stress. When *Normal Processing Rate* equaled to 0.96, the *Event Processing Rate* did not drop. This was because the initial stress level was just slightly increased (Figure 3-18), and the positive effect of stress boosted the processing rate. In this case, the product of positive and negative effects of stress was 1.09*0.9554 = 1.041 (dimensionless) (Section 3.3.3). In other words, under slightly higher stress, people could perform the tasks faster than normal. Such behaviors are also influenced by *Positive Effect of Stress* and *Negative Effect of Stress*.

In the next step, the Drained from Boredom loop was added for testing. Three scenarios are compared: 1) without the Drained from Boredom loop; 2) with the Drained from Boredom loop and *Normal Processing Rate* equals 2 tasks/minute; 3) with the Drained from Boredom loop and *Normal Processing Rate* equals 20 tasks/minute. The *Executive Control Resource* should decrease over time under low task load. With a much higher *Normal Processing Rate*, human operators would process the tasks faster and have more idle time. In such case, the level of boredom will be higher, causing a faster depletion of *Executive Control Resource*. This is reflected in the model outputs shown in Figure 3-19.

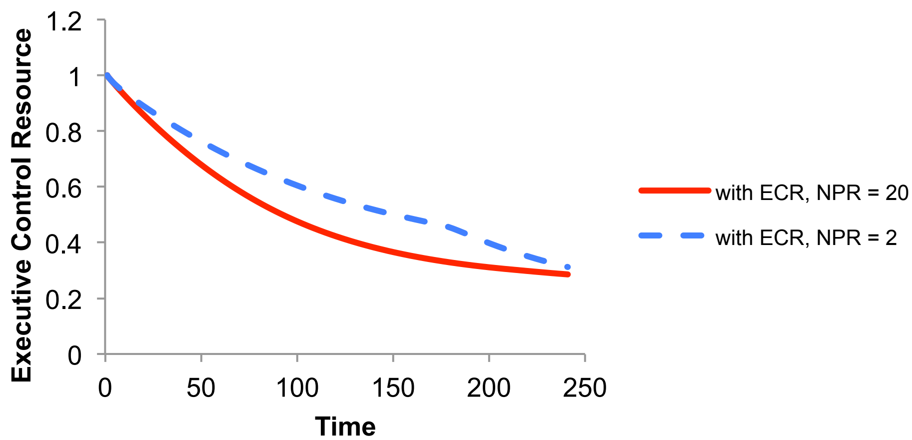


Figure 3-19: Change of Executive Control Resource

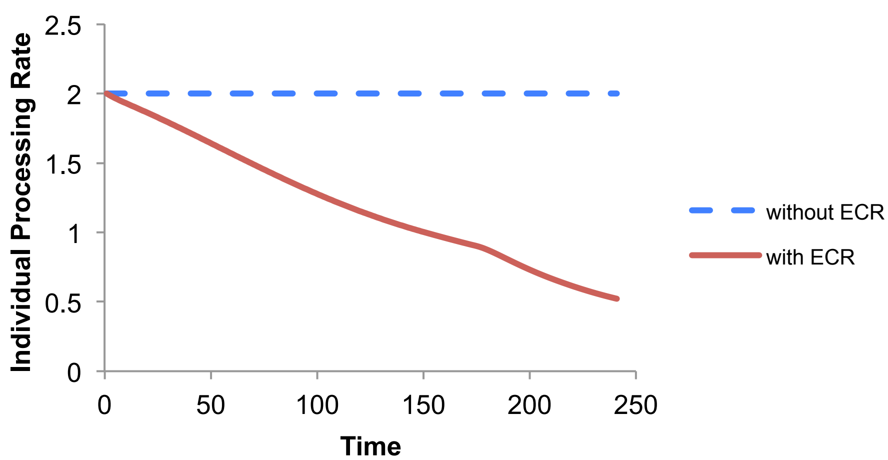


Figure 3-20: Impact of ECR on Individual Processing Rate

The level of *Executive Control Resource* has an impact on task processing and performance. As shown in Figure 3-20, without the Drained from Boredom loop, *Individual Processing Rate* stays at a constant level, which equals *Normal Processing Rate*. With a decreasing *Executive Control Resource* level, *Individual Processing Rate* is also decreasing.

Finally, the Attention Control loop was added into the model for testing. In Figure 3-21, it shows that *Attention on Task* decreases earlier and faster when the human operator processes the tasks at a higher rate. Intuitively, when the processing rate is higher, human operators have more idle time, leading to decreased *Executive Control Resource* and *Attention on Task*.

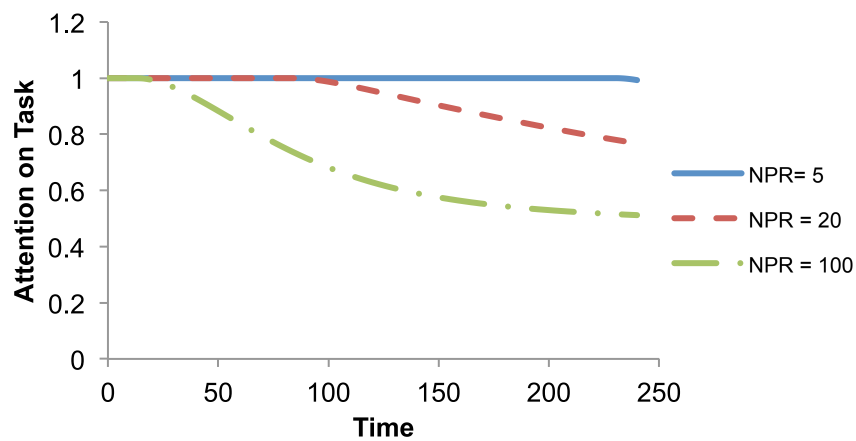


Figure 3-21: Change of Attention on Task

3.4.4 Extreme Condition Tests

Extreme condition tests inspect the robustness of the model in extreme conditions. The model should behave in a realistic fashion no matter how extreme the inputs or policies imposed on it may be. Instead of using a continuous event arrival process as in the previous section, discrete event arrival processes are tested here as they allow for completely idle time when no event arrives and surge of peak time when several events arrive together. The model behavior under both extremely low task load and high task load can then be observed. The behavior during the transition from low task load to high task load can also then be examined.

The total number of tasks is kept the same in all cases, which are 240 tasks within 240 minutes. The *Normal Processing Rate* is set to 100 tasks/min. As shown in Figure 3-22, *Attention on Task* decreases over time when the tasks arrive continuously at the rate of 1 task/minute. This is because the processing rate is much higher than the task arrival rate at each time step. When a batch of 20 tasks arrives every 20 minutes, *Attention on Task* still decreases but with fluctuations. 20 tasks are still manageable, but the stress level is higher than the continuous arrival scenario. When a batch of 240 tasks all arrive at one time, *Attention on Task* first decreases due to idleness, then increases to maximum level with the high task demand.

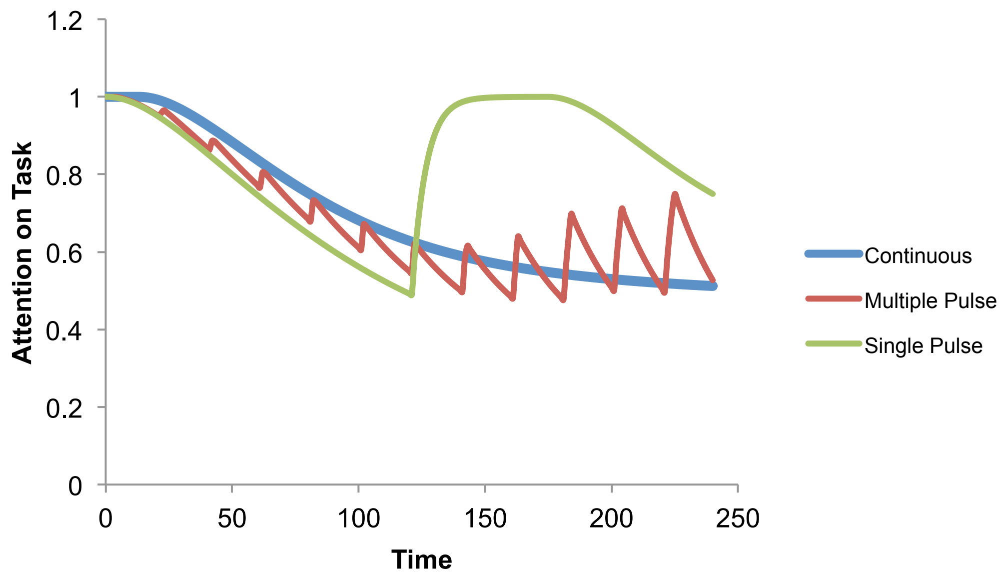


Figure 3-22: Attention on Task with Different Arrival Processes

Another extreme condition test is to observe the system behavior when *Executive Control Resource* and *Attention on Task* are initialed at zero. The model was tested by initializing either *ECR* or *Attention on Task* or both at zero. In both cases, human operators would not be able to process the tasks, resulting in an *Event Processing Rate* equaling to zero in the model outputs. These tests show that the model represents realistic human behavior even under extreme conditions.

3.5 Chapter Summary

In summary, a System Dynamics model of human attention and performance in long duration, low task load scenarios was created. Three dynamic hypotheses regarding the change of attention and performance in such task environments were presented. The modeling process was described. The final version of the model was introduced with five modules, namely task characteristics, task processing and performance, stress, workload, executive control, and attention management. Six feedback loops that generate the dynamic behaviors are described. This model was built based on previous research in related fields. Model boundary, dimension consistency, model structure, and behavior under extreme conditions were inspected. In the next three chapters, this model is tested using three experiment data sets collected in different types of tasks.

4 Hypothesis 1: Modeling the Impact of Hours of Boredom

This chapter describes a human subject experiment data set that was used to test dynamic hypothesis 1, which states that individuals reduce their attention on primary tasks under low task load due to the reduction in executive control. The task, experiment design and key results are described. Parameters used in the model are presented. A comparison between model outputs and experiment results evaluates the ability of the PAL Model to replicate attention and performance decrease under low task load. In order to evaluate the ability of the model to predict the impact of changes in system design, a second set of experimental data collected with the same testbed is used. A system improvement approach named 'increase task engagement' is evaluated using the PAL model, which demonstrates that the model can be used to facilitate system design.

4.1 Experiment Description

This section describes the long duration, low task load human performance experiment in which 30 participants endured a 4-hour experimental session acting as operators engaged in supervisory control of networked autonomous vehicles for a search and track mission. The task, experimental design, and results are discussed.

4.1.1 Task

This experiment was conducted using the Onboard Planning System for Unmanned vehicles Supporting Expeditionary Reconnaissance and Surveillance (OPS-USERS) test bed (Hart 2010). The simulation allowed a single operator to supervise multiple autonomous vehicles in a search and track mission. The operator was assisted by an automated planner for scheduling the vehicles' tasks. The operator interacted with the automated planner via a decision support tool to alter automation-generated schedules and approve desired plans.

The objective of the task was to command four heterogeneous vehicles to search the area of responsibility for hidden targets and then keep tracking the targets upon finding them. Once a target was found, the user identified the target as hostile, unknown, or friendly, and assigned a priority level to it. To track the positions and movements of hostile and unknown targets, one or more vehicles continually revisited them as often as possible. All the searching and tracking are scheduled by the auto-planner and can be executed automatically.

The operators spent most of the mission time monitoring the system, and interacted with the interface only occasionally to respond to replanning requests from the auto-planner to update the task schedule or create search tasks as requested by the commend center. In addition to the requested interaction, the operators could also create extra search tasks on unsearched locations via a map for the vehicles to explore or initiate extra replanning.

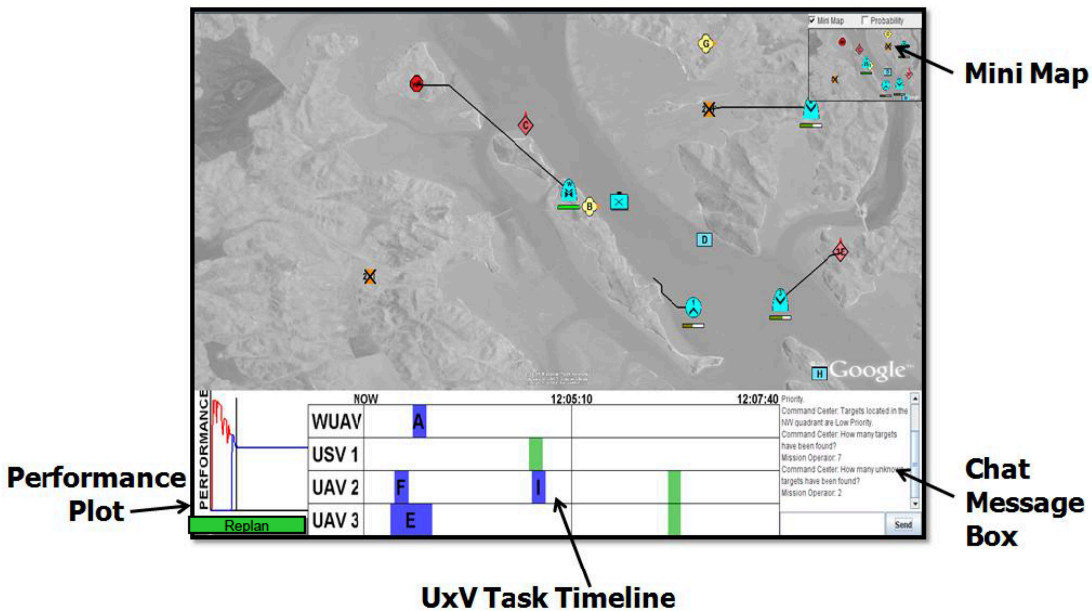


Figure 4-1: Map Display

Figure 4-1 shows the top layer display of the user interface of OPS-USERS, called the Map Display. It shows symbols representing the vehicles, search tasks, loiter tasks, and targets. The upper right-hand corner of the Map Display is a mini map, which provides an overview of the map. The timeline at the bottom gives temporal event information for the next five minutes for the four vehicles. Green bars indicate times of refueling. Blue bars indicate times of performing a task with a letter indicating the task type. White space indicates vehicle idle time or travel time between tasks. The lower left-hand corner of the Map Display is a performance plot, which shows the predicted performance score and the actual score of the human-automation system over time. When the actual score is lower than the predicted score, the auto-planner prompts the operator to accept the proposed task schedule plan in order to improve performance. In order to accept the new task schedule, the operator needs to click the green "Replan" button to enter another view, which is presented in Figure 4-4 and explained later. The command center sends intelligence information to the operator via the chat message box located in the lower right-hand corner of the Map Display.

The main tasks for the operator include creating/editing/deleting search tasks, identifying targets and replanning. A primary mission objective is to search unknown areas to look for targets. The vehicles automatically search the area of interest using their own onboard computer search algorithm using a consensus search method (Cummings et al. 2012). The operator can create a search task at a particular location by right clicking the location on the map. A search task creation window as shown in Figure 4-2 pops up, allowing the operator to assign a priority level and a time window for the search task. The operator can also edit an existing search task by right clicking it, which will bring up the same window. Operators create search tasks when they receive requests from the command center in the Chat Message Box. They can also create extra search tasks on the map in addition to the ones requested in an effort to improve search performance.

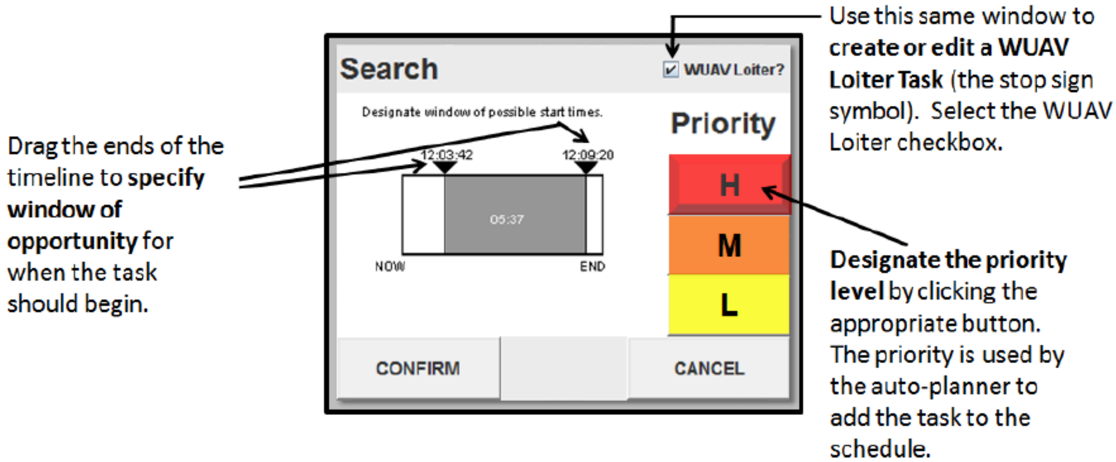


Figure 4-2: Search Task Creation Window

In OPS-USERS simulation, it is assumed the vehicles have automatic target detection capability. The target identification window pops up automatically when one of the vehicles discovers a target. In the experiment, the target identification task was simplified to recognizing the target symbols instead of the actual images. The operator need to pan through the target identification window to find the target symbol, classify it as hostile, unknown, or friendly based on its shape and color, and designate a priority level using intelligence information from the chat message box. Figure 4-3 shows the process of target identification.

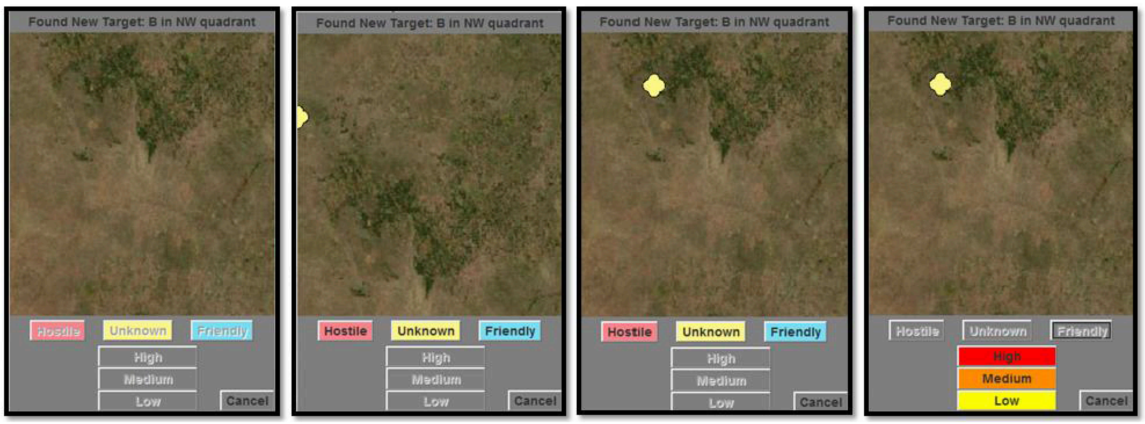


Figure 4-3: Target Identification Window

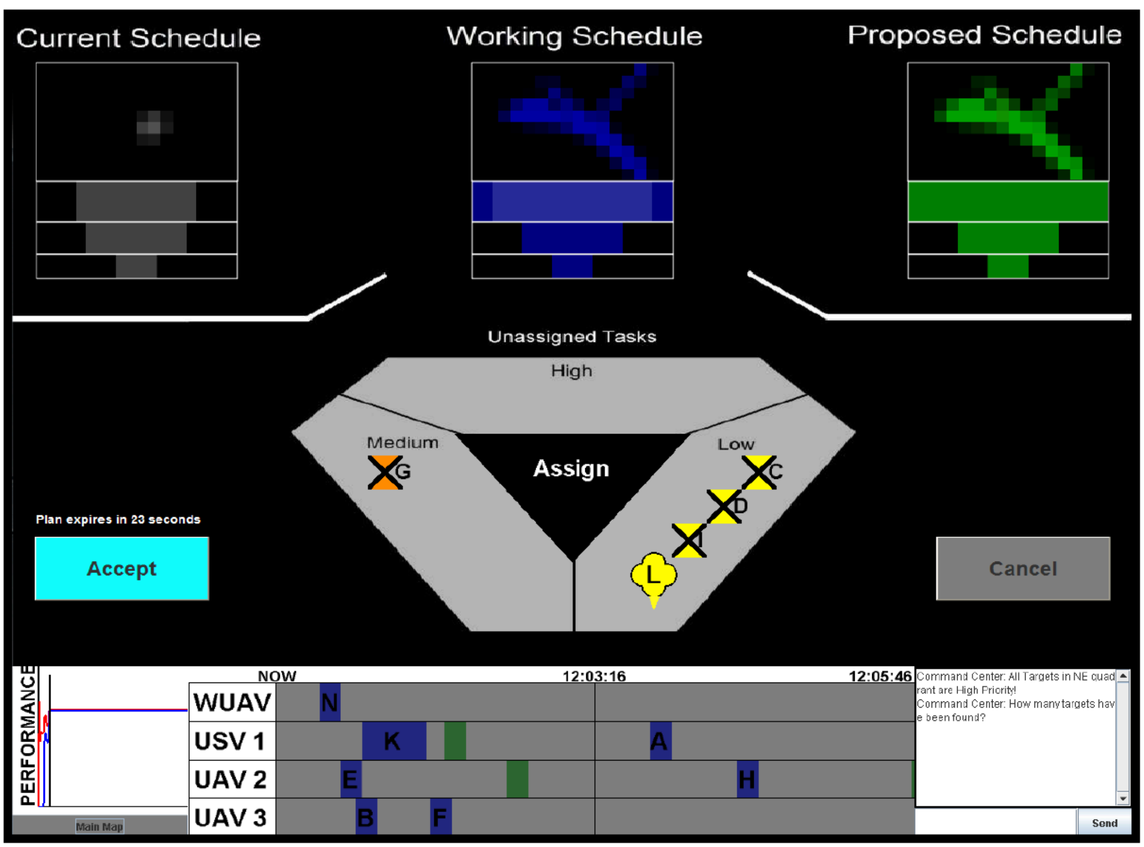


Figure 4-4: Schedule Comparison Tool

The tasks are scheduled by an automated planner, which can be updated via a decision support tool called the Schedule Comparison Tool (SCT). The green "Replan" button at the bottom left corner of the Map Display allows the user to view the interface of SCT, which is shown in Figure 4-4. Replanning is often suggested by the automation when new tasks are created, task priorities are changed, or there is a potential to improve performance. Human operators can also initiate extra replanning to update the schedule. All the search and track tasks are scheduled and assigned to the vehicles via the SCT. Operators can request the automation to change the assignment of a task. The three geometrical forms at the top of the SCT are configural displays showing the potential performance of three schedules. Each configural display shows the map area that will be covered, and the percentages of high, medium, and low priority tasks to be completed for a given schedule. The dark gray form on the left is the current schedule being carried out by the vehicles. The green form on the right is the proposed schedule from the automated planner. The blue schedule in the center shows the working schedule resulting from a collaborative effort between the human operator and automated planner. This collaborative approach has been shown to improve operator performance and situational awareness comparing to a fully automated approach (Clare 2013).

To create a long duration, low task load scenario, the speed of the vehicles was greatly reduced. It took almost an hour for a vehicle to move from one side of the map to the other. There were only 4 hidden targets to find in the 4-hour mission. To maintain low operator task load throughout the entire session, the 4 targets could not be found all at once. One of the 4 targets was "uncloaked" at the beginning of each hour. Even if an operator was able to use his or her vehicles to search the entire map area within the first hour, only one target would be found and identified, leaving the other 3 targets hidden until their future "uncloaking" times. The participants were unaware of this uncloaking activity. Moreover, the participants were prompted to replan only once every 10 minutes, 20 minutes, or 30 minutes. However, participants were allowed to create additional search tasks or initiate extra replanning if they chose to do so.

4.1.2 Experiment Design

The independent variable for this experiment was the replan interval, or the rate of how often the participant was prompted to collaborate with the automation to update the task schedule. Each participant was given a fixed replan interval of 10 minutes, 20 minutes, or 30 minutes. The 30-minute replan interval was designed to produce operator utilizations around 5%; the 20-minute replan interval was predicted to result in operator utilizations close to 10%; and the 10-minute replan interval was designed to place operator utilization at 15%.

Demographic data of the participants were collected using a pre-experiment survey. Participants were videotaped during the test session to capture their behaviors throughout the study. Workload and performance metrics were collected automatically by the simulation during the test session. Thirty minutes prior to the end of the simulation, the timeline grayed-out, indicating that there were no future events as the simulation came to a close. After the test session, participants filled out a survey, rating their busyness level, confidence in the actions they took, and subjective self-rated performance. They also indicated whether they were distracted or not, and listed any distractions they encountered during the test session.

The dependent variables include objective workload, objective performance metrics, subjective self-rated performance metrics, and attention state metrics obtained via video data. Objective performance metrics include mission effectiveness of searching and tracking, human behavior efficiency as measured by reaction time to system prompts, and human automation collaboration metrics as measured by extra operator-driven interactions with the automation. However, only the searching performance is used for comparison with the system dynamics model, as it was the primary task performance.

Potential distraction sources were available to the participants during the experiment, such as Internet access via a secondary monitor, magazines, refreshments, cell phones or books. Thirty participants were tested in groups of three, resulting in potential distraction from social interactions. Video data was coded to measure the participants' attention states during the experiment test session. Each participant's time was classified into percentage of time spent in: (1) directed attention, or appearing focused on the interface; (2) divided attention, or multitasking while still paying attention to the interface; and (3) distracted attention, or doing anything other than monitoring or interacting with the simulation interface. The detailed criteria for video coding are as follows:

1) Directed Attention

The participant appears focused and is only monitoring or interacting with the interface and not doing any other task.

2) Divided Attention

The participant has eyes on the interface screen, but multitasks by eating, stretching, talking, playing Minesweeper, etc.

3) Distracted Attention

The participant is not paying attention to the interface at all. Examples include: sleeping, eating a meal without looking at the interface, discussions with participants' backs turned to the computer, reading a book, etc.

4.1.3 Results

A thorough analysis of the experiment data can be found in Hart (2010). For the purpose of model testing, key results related to utilization, performance and attention states are presented in this section.

Even though task demand was varied via the replan interval independent variable, the total utilization was not statistically different across the three replan intervals when tested using the Kruskal-Wallis test ($\chi_2 = 0.135$, $p = 0.935$). The average utilization of all participants was 11.4% with a standard deviation of 3.36%. A deeper investigation breaks utilization into two parts: required utilization and self-imposed utilization. Required utilization is the percentage of time a participant was required to spend interacting with the simulation, based on replan interval, number of search tasks created as prompted by the command center and number of targets found that required identification. Each participant's required utilization was specific to the replan interval independent variable, but varied due to the slightly different situation each one had in performing the tasks. The required interaction with the system was captured by the flow of *Event Arrival Rate* in the PAL model. In contrast, self-imposed utilization is the percentage of time a participant interacted with the interface by doing activities that were not required by the mission, such as extra replanning, creating participant-generated search tasks, and additional uses of the target identification window for editing target designations. These additional interactions were modeled as *Self-Imposed Event Arrival Rate* in the PAL model.

The data for each level of replan interval is presented in Table 4-1. The high level of self-imposed utilization shows that participants interacted with the system more than the system required on purpose, likely to combat the boredom purposefully induced. This added utilization may be due to the extra cognitive capacity that the participants had during the low workload scenario. Humans do not operate at low task load comfortably and the participants likely chose to increase task engagement to sustain their attention.

Table 4-1: Utilization by Replan Interval

Replan Interval	Required Utilization		Self-Imposed Utilization		Total Utilization	
	mean	S.D.	mean	S.D.	mean	S.D.
10 minute	2.41%	0.46%	9.04%	4.53%	11.44%	4.63%
20 minute	1.69%	0.14%	9.87%	3.65%	11.56%	3.66%
30 minute	1.58%	0.36%	9.59%	1.51%	11.17%	1.57%

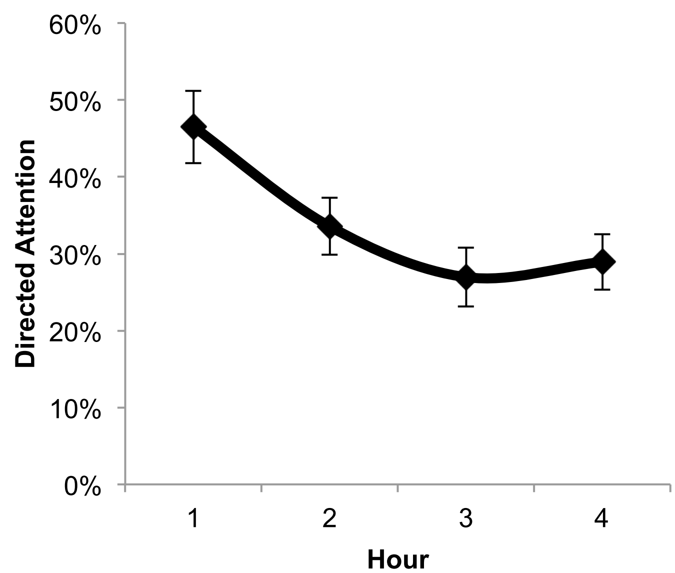


Figure 4-5: Changes of Directed Attention over Time

Analysis on attention states shows that participants spent an average of 34% (S.D. = 15%) of their time in a directed attention state, 22% (S.D. = 13%) of their time in a divided attention state, and 44% (S.D. = 20%) of their time distracted. Directed attention was also found to decrease over time. Figure 4-5 shows the average percentage of time in the directed attention state across all participants, as well as the standard error. A Repeated Measures

General Linear Model showed a significant difference in directed attention across hour intervals ($F = 21.953$, $p < 0.001$). Although the amount of directed attention is higher in hour 4 as compared to hour 3, the difference is not significant.

Directed attention is also shown to correlate with task engagement as reflected by operator utilization and extra interactions with the system in addition to required replans and search tasks created as requested by the system (extra search tasks and extra replans). Directed attention is moderately correlated with total utilization (Pearson's $\rho = 0.434$, $p = 0.017$). Total directed attention also is correlated with extra search tasks ($\rho = 0.509$, $p = 0.004$) and extra replans ($\rho = 0.580$, $p = 0.001$). Total divided attention correlated with extra search tasks ($\rho = 0.453$, $p = 0.012$) and extra replans ($\rho = 0.374$, $p = 0.042$). These results show that operators who had high levels of attention devoted to the primary task were also more engaged with the task. They initiated more interactions with the system in addition to the requested ones. Oppositely, total distraction correlated negatively with extra search tasks ($\rho = -0.684$, $p < 0.001$) and extra replans ($\rho = -0.689$, $p < 0.001$).

Consistently, self-imposed utilization (percentage of time spent on extra interactions) was found to correlate negatively with total distraction ($r = -0.406$, $p = 0.026$). This means that people who were distracted also had a lower task engagement level. These correlations show that attention state affects behaviors that comprise utilization. Although the causal relation between task engagement and attention is not clear with the current experiment data, it is reasonable to assume a feedback relation between the two variables. High attention level may motivate the operator to increase task engagement, which helps to maintain the attention level on the primary task in a low task load environment. These relations were captured in the PAL model via the Increase Task Engagement loop. If the operators' attention on the primary task is higher than the amount needed for the required interactions, they may choose to generate self-imposed events to utilize the spare attention.

Search performance is measured by an aggregated target finding score, which takes into account both the number of targets found and the time required to find each target. Total utilization was found to be a significant predictor of search performance ($\beta = -4.282$, p =

0.007) in a regression analysis. Since total utilization also positively correlates with directed attention, this shows that in long duration, low task load environments, directed attention may be increased with a higher level of task engagement, which could result in better performance.

4.2 Model Parameters

In order to test whether the system dynamics model in Chapter 3 could capture the change of human attention and performance in low task automated environments, the model outputs were compared with the experiment data. In the OPS-USERS scenario described previously, human operators did not interact with the vehicles directly to perform the tasks. Humans interact with the automated scheduling algorithm to improve the system performance. In the model, these interaction events represent tasks for the human. Additional model structure was added to reflect the impacts of such interactions on search performance, as shown in Figure 4-6. This is a change specific to this task environment. Total number of *Self-Imposed Events* is modeled as a stock with an inflow *Self-Imposed Event Rate*.

$$Self\ Imposed\ Event\ Rate(t) = Baseline\ Self\ Imposed\ Event\ Rate * MAX(-Attention\ Gap(t), 0) \quad (27)$$

$$Self\ Imposed\ Event(t) = \int_t Self\ Imposed\ Event\ Rate(t)\, dt \quad (28)$$

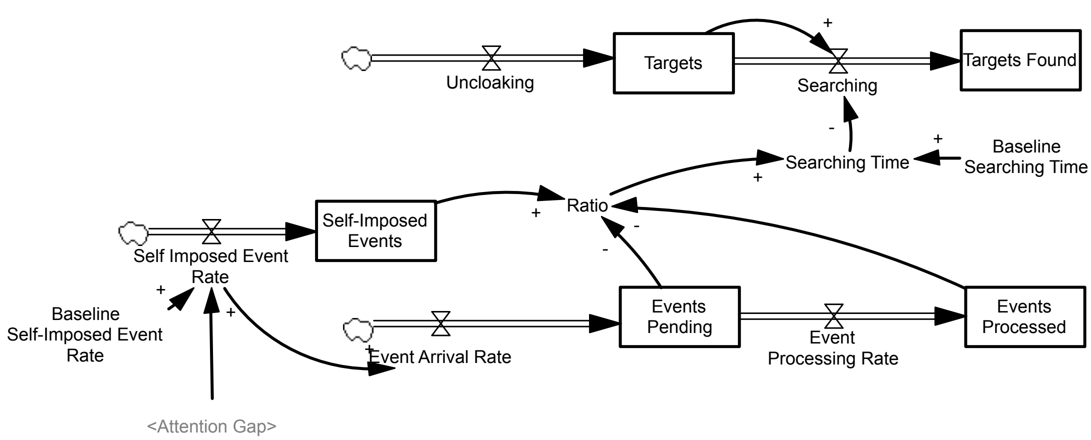


Figure 4-6: Model Structure for Searching Performance

Another chain of stocks and flows captures the target uncloaking as an event arrival process and searching process. The search process is influenced by the total number of targets available to be searched, and the time to find a target, called *Searching Time*. *Searching Time* is influenced by human interactions with the system. Specifically, the more interactions completed by the human operator, the shorter the *Searching Time*. In other words, the targets will be found faster if the human operator creates more search tasks or updates the task schedule adequately via replans. This is supported by experimental results that total utilization of operator correlates positively with performance. In addition, the higher the percentage of *Self-Imposed Events* in total number of interactions is, the shorter the *Searching Time*. This is supported by another study using the same testbed with high task load. It was shown that self-imposed interactions could improve the performance of human-automation collaboration (Clare 2013). These relationships are captured in Equation (29)-(32).

$$Targets(t) = \int_t \left[Uncloaking(t) - Searching(t)\right] dt \tag{29}$$

$$Targets\ Found(t) = \int_t Searching(t)\, dt \tag{30}$$

$$Searching(t) = \frac{Targets(t)}{Searching\ Time(t)} \tag{31}$$

$$\begin{aligned} &Searching\ Time(t) \\ &= Baseline\ Searching\ Time * \frac{Events\ Processed(t)}{Events\ Pending(t) + Events\ Processed(t)} \\ &\quad * \frac{Self\ Imposed\ Events(t)}{Events\ Pending(t) + Events\ Processed(t)} \end{aligned} \tag{32}$$

The parameters used in the model are shown in Table 4-2. *Baseline Self-Imposed Event Rate* is set to 1.5 tasks/minute, which equals the average self-imposed event rate in the experiment data. The average number of total extra interactions was 369.53. Dividing this number by the length of mission, 240 minutes, gave us 1.53 tasks/minute. *Uncloaking* of targets happened at around 5 minutes, 33 minutes, 125 minutes and 184 minutes into the experiment. This is captured using four discrete events by using PULSE function in Vensim®. In the experiment data, since not all the targets were found, it was unable to estimate the real *Searching Time*. Estimating the search time based only on the targets found would be an underestimation of the real *Searching Time*. In the model, *Baseline Searching Time* is set to 121.98 minute using the model calibration function in Vensim®.

Table 4-2: Model Parameters

Model Parameters	Values	Parameters for Nonlinear Relationships
Baseline Self-Imposed Event Rate	1.5 task/minute	$k_1 = 1$
Uncloaking	PULSE (5, 1)+PULSE (33, 1)+PULSE (125, 1)+PULSE (184, 1)	$k_2 = 10$
Baseline Searching Time	121.98 minute	$c_1 = 1.69$
Exogenous Event Rate	PULSE TRAIN(0, 1, 20, 240)	$c_2 = 1.44$
Normal Processing Rate	20 tasks/minute	$k_3 = 4.22$
Time Window	5 min	$k_4 = 2.09$
Depletion Time	240 min	$k_5 = 1.25$
Restore Time	240 min	$c_5 = 1$
Executive Control Resource: Initial value Maximum Level Minimum Level	 0.5 0.5 0	$m_5 = 0.51$
Attention on Task, Initial value	0.5	$k_6 = 2.07$
Average Time to Distract	25 min	$c_6 = 3.33$
Time to Refocus	6 min	$k_7 = 1$
End of Mission Effect	30 min	$c_7 = 1.03$

Exogenous Event Rate is set based on the replan interval in the experiment. Although there are three levels of replan intervals in the experiment, a single value of 20 minutes was used in the model. This is because the experiment data shows no significant difference on performance resulting from replan intervals. The PULSE TRAIN function in Vensim® is used to model the arrival of one replan event every 20 minutes. *Normal Processing Rate* equals 20 tasks/minute, because experiment data shows that each replan event needs 3 seconds to process on average. *Time Window* for calculating *Average Processing Rate* was chosen to be 5 minutes. This variable was used as the sampling time interval to calculate the exponential moving average of *Event Processing Rate*. Using 5 minutes reflected recent changes in workload.

For *Executive Control Resource*, both the *Depletion Time* and *Restore Time* are set to 240 minutes, which is the length of the mission in the experiment. The minimum level of *Executive Control Resource* equals zero. The initial value and the maximum level of *Executive Control Resource* are both set to 0.5 to be consistent with the initial attention level of 0.5. Attention on Task is initialized at 0.5, because the percentage of time in the state of directed attention is about 50% at the beginning of the experiment. *Average Distraction Time* is 25 minutes because previous research shows that sustained attention decreases around 20-30 minutes into the mission (Mackworth 1957). A survey for general workplace shows that an average of 9.28 minutes was needed to refocus on a task after being interrupted (Russ and Crews 2014). In another experiment that will be described in Chapter 5, it took 6 minutes on average to increase the attention on task when an emergency event happened (Thornburg et al. 2011). Thus, the refocus time is set to 6 minutes in the model. Thirty minutes prior to the end of the simulation, the timeline grayed-out, indicating that no future events were visible as the simulation came to a close. This may affect the attention state of human operators, as shown in the slight increase of attention at the end of the experiment. An *End of Mission Effect* is added to the model to reflect this change.

There are other parameters used in the model to describe the nonlinear relationships as presented in Chapter 3. The values of these parameters are chosen using the model calibration function in Vensim®. The visualization of the nonlinear relationships is included in Appendix A. The calibration process is described in Appendix D.

4.3 Behavior Reproduction Test

One key aspect of model testing is to assess the model's ability to reproduce the behavior of a system, which is called the behavior reproduction test (Sterman 2000). In order to do this, two key outputs of the model are compared with the experiment data. A quantitative assessment of the model's fit to experimental data is provided in Table 4-3. The model-data variation was decomposed by splitting the Mean Square Error (MSE) into three components: bias (U^M), unequal variation (U^S), and unequal covariation (U^C). The ultimate goal of a model fit is to have small errors between the model and data (indicated by [Mean Absolute Percent Error] MAPE, [Root Mean Square Error] RMSE, etc.), with most of the error due to unsystematic, or random, variation (concentrated in U^C) (Sterman 1984, 2000). Overall, a high level of R^2, low level of MSE, and high level of U^C indicate a good model fit.

Attention on Task is compared with the average level of directed attention in the experiment data. Percentage of time in the state of directed attention is summarized by minute, as shown in the thin blue line in Figure 4-7. The thick red line shows the change of *Attention on Task* in the system dynamics model. As can be seen from the graph, the model output is well aligned with the experiment results, demonstrating the ability of the model to capture real human behavior in a low task load scenario.

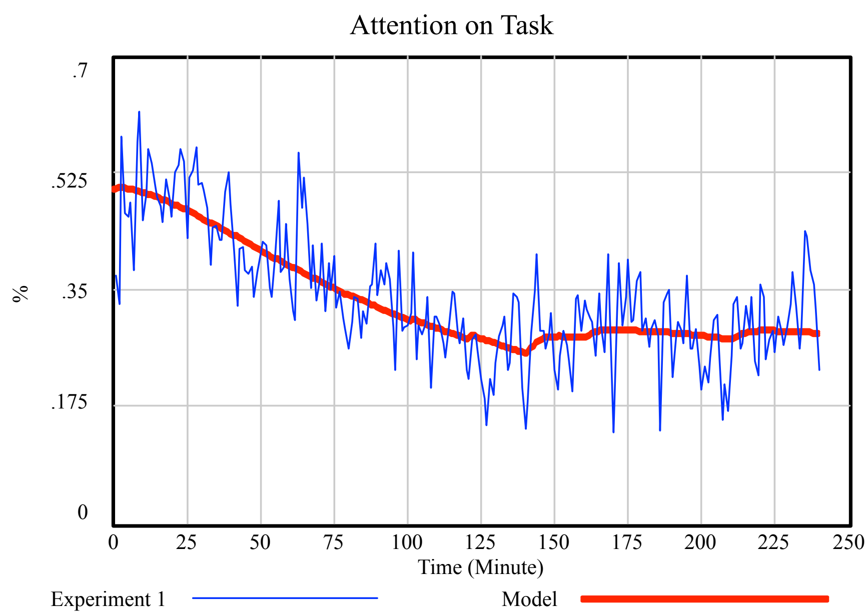


Figure 4-7: Comparison of Attention on Task

Table 4-3: Simulation to Experimental Data Fit Statistics (without Alerts)

Summary Statistics	Attention on Task	Targets Found
Coefficient of Determination (R^2)	0.627	0.983
Root Mean Square Error (RMSE)	0.058	0.118
Mean Square Error (MSE)	0.003	0.014
Bias component of MSE (U^M)	0.002	0.027
Variation component of MSE (U^S)	0.126	0.593
Covariation component of MSE (U^C)	0.872	0.381

As presented in Table 4-3, the model has a good fit to the experimental data with coefficient of determination (R^2) values of 0.627. The largest component of MSE is U^C, which equals 0.872. This means that most of the error was due to unsystematic, or random, variation, indicating a good model fit. Dynamic hypothesis 1, which says individuals reduce their attention on primary task under low task load on average, is well supported by both experimental evidence and model outputs. The curve of attention change from the model output is much smoother than the experiment data. This is because human attention is affected by environmental cues, which are outside the model boundary. However, since the purpose of the model is to capture the attention change in a long duration mission, these variations are less important comparing to the general decreasing trend.

The change of *Targets Found*, a primary performance metric, is also compared with the average search performance in the experiment data, as shown in Figure 4-8. The model has a good fit to the experimental data with coefficient of determination (R^2 = 0.983). The largest component of MSE is U^S instead of U^C, which equals 0.593. This shows that the model has unequal variation compared to the experiment data. U^M is at a low level of 0.027, which is desired. Although the value of U^C is a bit low, it is clear that the PAL Model can capture human performance in this task scenario.

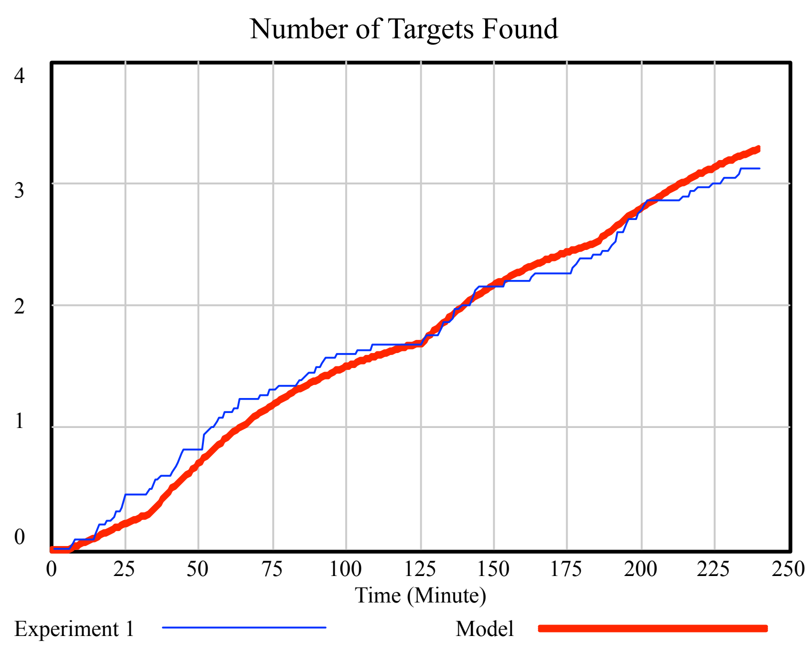


Figure 4-8: Comparison of Search Performance

4.4 System Improvement Test: Effect of Attention Alert

One purpose of building the model is to use it to facilitate the system design process. Ideally, changes to a system can be tested in the model first to save time and reduce cost. In this section, the ability of the model to evaluate system changes is assessed.

As suggested in the experiment and the model, directed attention plays an important role in human operators' performance in low task load automated environments. One possible way to improve directed attention is simply giving attention alerts to the human operators over the course of mission. A second study using the same test bed was conducted to examine whether cyclical automated alerts used to remind subjects to assess the quality of the automation's performance could improve overall performance (Mkrtchyan et al. 2012). The alerts were implemented in the form of auditory alerts that were pre-programmed in the interface. The alerts were designed to be distinct from all the existing aural alerts within the interface. The alerts consisted of four distinct chimes approximately 300ms long that resembled a doorbell sound. Between the first two and last two chimes there was a 400 ms pause. Between the second and the third chimes the duration of the pause was 1.2 seconds. All participants were required to wear wireless headphones at all times to hear the alerts. Twenty-four participants completed the tasks either with or without the attention alerts. The results show that the alerts had a marginal effect ($p = .052$) for the amount of area searched by the vehicles, but did not improve the number of targets found. Directed attention was slightly higher with the attention alerts, but the difference was not significant ($t(8) = 0.71$, $p = 0.49$).

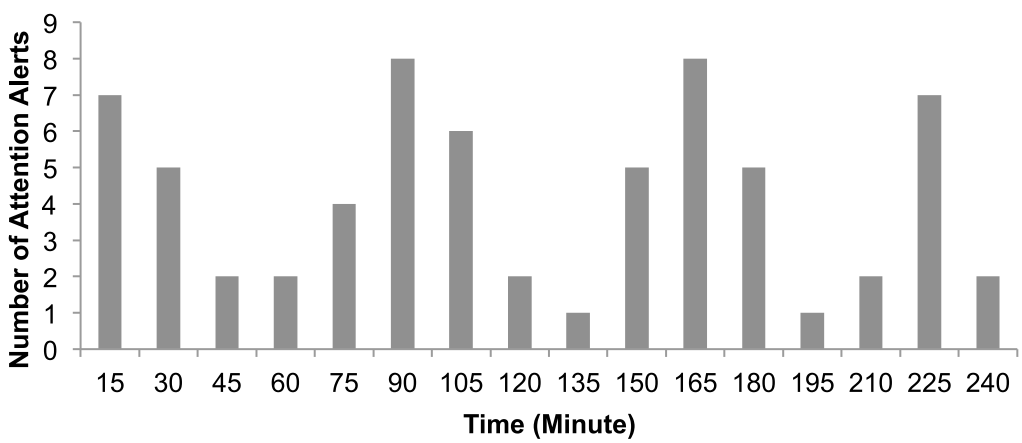


Figure 4-9: Number of Attention Alerts in Each 15-Minute Block

The effectiveness of attention alerts can also be assessed using the system dynamics model. Two changes were made to the model. First, there were six targets to be found in the second experiment. Second, it was hypothesized that the attention alerts would impact the *Intended*

Allocation of Attention on Task. Whenever an attention alert was given, the human operators would increase their level of directed attention. The timing of the attention alerts was set based on an experiment setting. The number of attention alerts in each 15-minute block used in the experiment is presented in Figure 4-9. In total, there were 67 alerts during 240 minutes in the experiment. While these attention alerts were not evenly distributed across time in the experiment, an average value of one alert every 4 minutes was used in the model to approximate the total number of alerts in the experiment. While the impact of each alert on attention was unknown, a coefficient of 0.4 was assumed in the model. The impact of attention alerts is modeled in Equation (33) and (34).

$$\begin{aligned} &Indicated\ Demand\ for\ Attention\ on\ Primary\ Task\ (t) \\ &\qquad = Min((1 + Effect\ of\ Stress(t) + 0.4 * Attention\ Alert(t)) \\ &\qquad\qquad * Executive\ Control\ Resource(t), 1) \end{aligned} \tag{33}$$

$$Attention\ Alert(t) = PULSE\ TRAIN(\ 0\ , 1, 4\ , 240\ ) \tag{34}$$

Based on Equation (33), the percentage of change on attention with the alert is calculated as:

$$\begin{aligned} &\frac{(1 + Effect\ of\ Stress(t) + 0.4 * 1) * Executive\ Control\ Resource(t)}{(1 + Effect\ of\ Stress(t) + 0.4 * 0) * Executive\ Control\ Resource(t)} - 1 \\ &\qquad = \frac{0.4}{1 + Effect\ of\ Stress(t)} \end{aligned} \tag{35}$$

Intuitively, the increase of attention is small when the operator is already stressed from the task load. The increase of attention is large when the operator is less stressed due to low task loading. Since the average value of *Effect of Stress* is 0.2 from a previous simulation run, attention alerting results in about a 30% increase on *Indicated Demand for Attention, as* compared to when no attention alert was given in the model. This is a reasonable assumption given the simplicity and short duration of the alerts. All the other parameters were kept the same as in the previous model.

The model outputs were compared with experiment results again as shown in Figure 4-10 and Figure 4-11. In Figure 4-10, the two thin lines are experiment data under two alert conditions, and the two thick lines are model outputs. For attention, the fit of the model output to the experiment data is not very good in each condition. Overall, the predicted level of attention is lower than the experiment data. However, comparing the model outputs under two alert conditions only, the level of attention is not improved much when attention alerts are presented. This is consistent with the experiment results that the increase on attention with alerts was not significant.

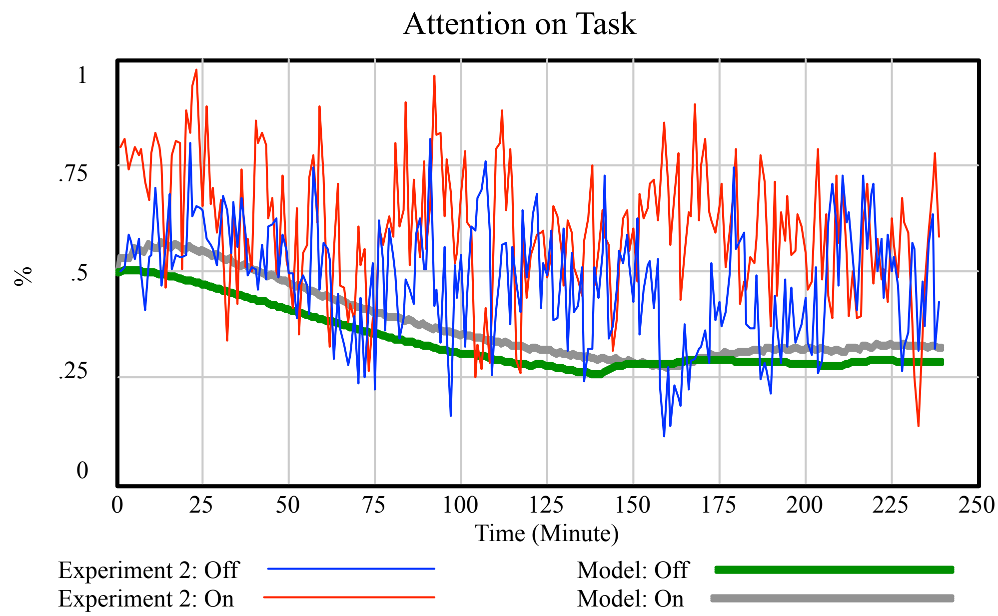


Figure 4-10: The Effect of Attention Alerts on Directed Attention

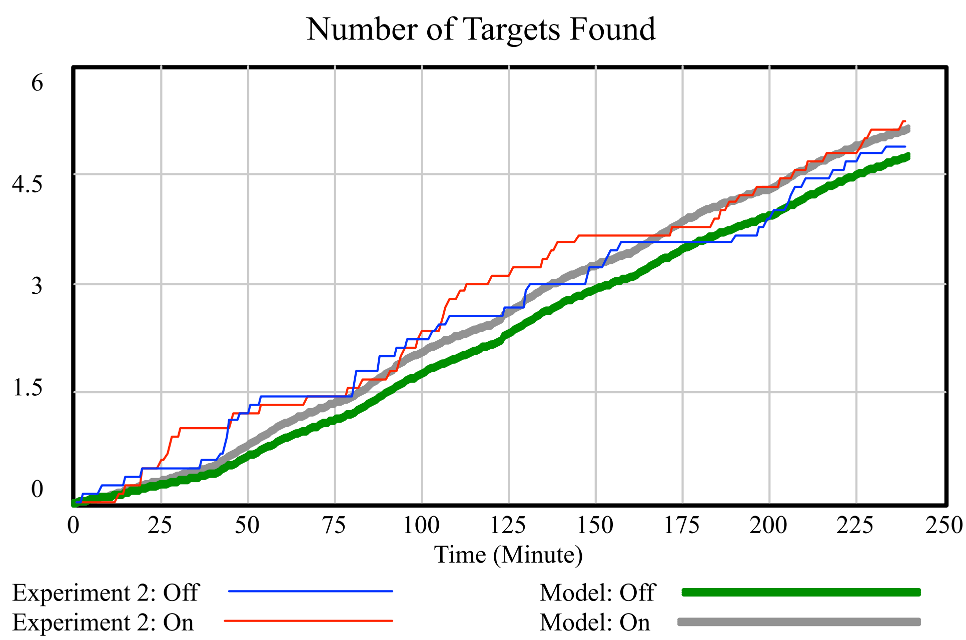


Figure 4-11: The Effect of Attention Alerts on Search Performance

Detailed statistics of the model prediction are provided in Table 4-4. R^2 values were negative for both alert conditions, indicating a poor fit. When examining the components of MSE, U^C was the largest component (0.582) when alerts were used, which was desired for a good model fit. The second largest component of MSE was U^M, which equals 0.347. This indicates that there is a bias in model prediction. For the condition with no attention alerts, U^C was the second largest component (0.379). A relatively large component of U^M (0.522) indicates that the model prediction was different from the experiment data with a bias. The poor fit of *Attention on Task* may also be influenced by different populations used in the two experiments, relatively small sample sizes, large variation of attention data, and data quality issues in the second experiment. However, the purpose of the model is to aid the relative comparison between different system design options. The impact of absolute prediction error is less severe in such cases.

Table 4-4: Simulation to Experimental Data Fit Statistics (with Alerts)

	With Alerts (On)		Without Alerts (On)	
Summary Statistics	Attention on Task	Targets Found	Attention on Task	Targets Found
Coefficient of Determination (R^2)	-0.517	0.967	-1.021	0.940
Root Mean Square Error (RMSE)	0.167	0.259	0.193	0.349
Mean Square Error (MSE)	0.028	0.067	0.037	0.122
Bias component of MSE (U^M)	0.347	0.002	0.522	0.636
Variation component of MSE (U^S)	0.071	0.481	0.099	0.035
Covariation component of MSE (U^C)	0.582	0.517	0.379	0.329

The model provides a much better fit for the performance data as shown in Figure 4-11. It also correctly reflects the fact that search performance was improved by the attention alerts only slightly, which was also observed in the experiment. For performance as measured by Targets Found, R^2 was 0.967 when there were alerts, and 0.940 when there were no alerts. The high R^2 indicates a good model fit. When examining the components of MSE, U^C was the largest component as desired for a good model fit under the condition with attention alerts, which equals 0.517. Under the condition without attention alerts, U^M was the largest component of MSE, which equals 0.636. This indicates a bias in model prediction. From Figure 4-11, it can be seen that model prediction was slightly lower than experiment result for performance.

In summary, the model predictions show that attention alerts were not very useful in increasing attention on primary task and system performance. From the system dynamics perspective, this is because the dominating loops of Attention Control and Drained from Boredom were not fundamentally changed by attention alerts. In other words, attention alerts were not sufficient to stop or even slow down the decrease of attention on primary task. Low task loading was still driving the decrease in attention. As a result, it does not make sense to use the alerts in the way proposed in the experiment. If the designers could have tested this mitigation strategy using the PAL model before running the experiment, the time and cost of running such long duration experiments could have been saved. Thus the PAL model could be used to aid in the design process by helping designers understand the effectiveness of proposed mitigation strategies.

4.5 System Improvement Prediction: Increase Task Engagement

One goal of building the PAL model is to facilitate system design. In this section, a system improvement approach named 'increase task engagement' was evaluated using the model. In the task scenario described in this chapter, the level of attention decreased over time due to the low task load. One mitigation strategy that could be proposed is to increase task engagement by increasing attention on the task and away from other activities (distractions, mind wandering), and thus improving overall performance. For example, in one air traffic control monitoring study, task engagement was increased by requiring the controller to click on each aircraft as it entered the airspace, which mitigated the vigilance decrement after the operators were sufficiently trained for the task (Pop et al. 2012). In our study, task engagement might be improved if the operator was required to report the state of the vehicles once in a while.

In the PAL model, this approach was modeled by adding an additional causal link between *Attention on Other Activities* and *Self-Imposed Event Rate*. It was assumed that half of the attention on other activities could be guided to increase task engagement by proper system design. Hence, *Self-Imposed Event Rate* was increased by 0.5**Attention on Other Activities***Baseline Self-Imposed Event Rate* comparing to the original case. The impact of this change on attention was shown in Figure 4-12. The impact on performance was shown in Figure 4-13. It can be seen that both attention and performance were improved if task engagement was increased.

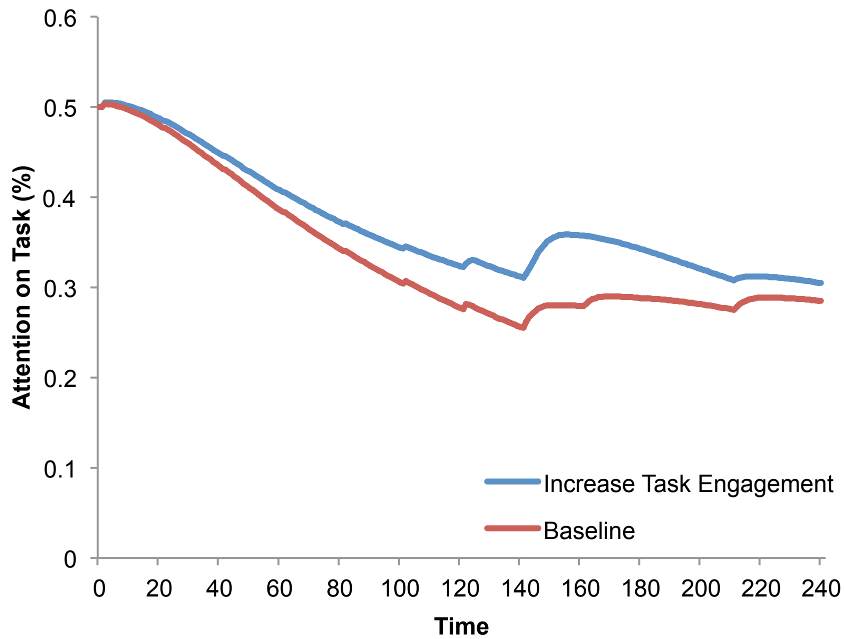


Figure 4-12: Impact of Increasing Task Engagement on Attention

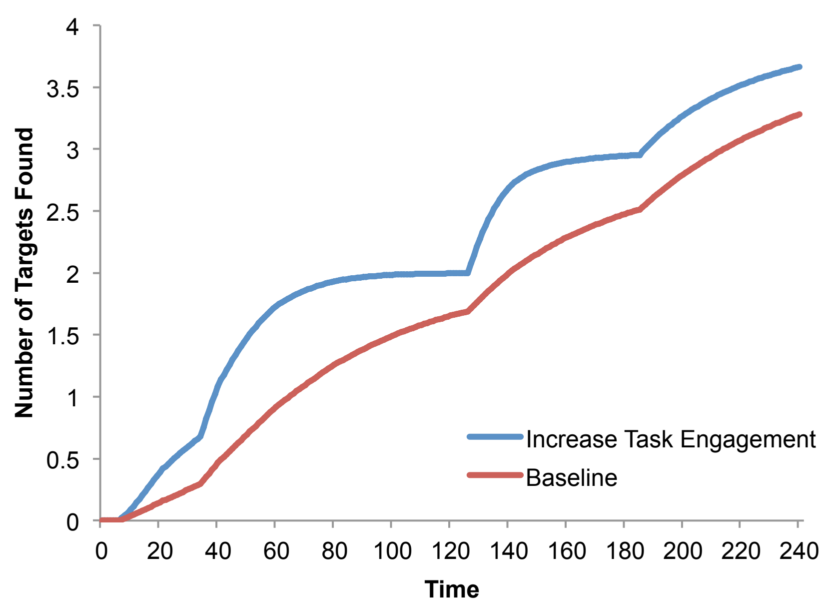


Figure 4-13: Impact of Increasing Task Engagement on Performance

While the results from the model predicted that guiding attention on other activities to increase task engagement is beneficial, detailed system designs cannot be directly derived from the model results. In practice, there might be several ways to increase task engagement, such as mindfulness training for the operators, setting a higher performance standard to motivate operators, etc. It requires careful examination of the specific task environment, and a combined effort of modeling and empirical studies to develop an improved system. However, the use of PAL can help guide designers in which improvements will likely have the greatest impact.

4.6 Chapter Summary

In this chapter, a long duration human subject experiment with low task load was introduced. In this experiment, participants controlled multiple vehicles to search for four targets with the aid of an automated planner in a four-hour mission. The system required only infrequent human interactions, each of which could be completed in seconds. Revisiting the task categories discussed in Chapter 3 (Table 3-1), this belongs to the last category in which both frequency and level of human effort are low. Two phenomenon were observed in the experiment. First, the level of directed attention decreased over time under low task load. This is captured mainly by the Drained from Boredom loop and Attention Control loop in the model. Second, people interacted with the system in addition to the system requirement. This is captured by the Increase Task Engagement loop in the model.

In order to assess the ability of the model to reproduce these behaviors and test dynamic hypothesis 1, model parameters were set based on experiment data, previous literature, and model calibration. The outputs of the model on attention and performance were compared with the experiment data. The comparison shows that the model could successfully replicate the experimental behavior in attention change and its impact on performance. As a result, dynamic hypothesis 1 was supported, which stated that individuals reduce their attention on primary task under low task load on average.

The model was further tested for its ability to predict the effect of system interventions. A second experiment data using the same test bed was used for this purpose. In this experiment, the effectiveness of attention alerts was evaluated. In the model, the parameters were unchanged except for slight changes to account for the difference in number of targets and the attention alerts. Comparison between the model outputs and experiment data shows that the model provided a good prediction on performance, but a less than ideal fit for attention data. However, in terms of evaluating the value of attention alerts relative to no attention alerts, the model predicts a slight increase on both attention and performance, but overall no significant effect, suggesting this design change was not worth exploring further. This is consistent with the experiment results. Another system improvement approach called 'increase task engagement' was evaluated using the PAL model. Model results predicted that this approach could improve attention management and performance.

Results of the behavior reproduction test and system improvement test are critical in building confidence in the model. In Chapters 5 and 6, the model is further tested using another two experiments in different task environments.

5 Hypothesis 2: Modeling the Impact of Moments of Terror

This chapter describes a human subject experiment data set that was used to test dynamic hypothesis 2. The task, experiment design and key results are described. Parameters used in the model are presented. A comparison between model outputs and experiment results is conducted to evaluate the ability of the model to capture the attention and performance in responding to an emergency event. In order to evaluate the ability of the model to predict the impact of changes in system design, part of the experiment data was used for model calibration and the rest was held for model prediction testing. The effect of restricting external distraction sources on attention management and performance in responding to an emergent event was evaluated using the model and compared with experiment results. The impact of a system improvement approach, secondary task, was predicted using the model.

5.1 Experiment Description

In long duration, low task load, safety-critical operating environments, human operators may need to switch from a passive monitoring state to an active alarm resolution state when an emergency event happens. In order to quickly identify the problem and respond with appropriate actions, high levels of attention and situation awareness are required. However, these high task load situations happen infrequently and are unpredictable. The long gaps in time between urgent or emergent events could lead to distraction, decreased cognitive ability and situation awareness.

This experiment investigated if the critical event onset time, the operating condition, or the attention state management has a significant correlation with operator performance (Thornburg et al. 2012). Thirty-six individuals participated in the experiment. This section introduces the task, experiment design, and the key results related to this modeling effort.

5.1.1 Task

A PC-based nuclear power plant control room simulator was used in this experiment to introduce tasks representative of actual nuclear power plant control. The interface of the simulation, called Human Operator Monitoring Emergent Reactors (HOMER), is shown in Figure 5-1. The primary task of HOMER users is to monitor the main interface and ensure that four mini reactors are functioning properly, including making sure no adverse events occur, making sure all parameters are in range, and responding to questions that appear in the popup window.

The HOMER is a basic representation of a digital 4-loop nuclear power plant control console. Each loop is displayed at each corner of the interface, showing the flow of coolant and steam from the central reactor vessel to the turbine and cooling tower. Within each loop, there are several components (pumps, generator, valves, gauges, etc.) that may or may not require user interaction that are referenced within the emergency procedure. Each gauge also displays the name of the measurement, the value of the measurement, whether the measurement is within limits or out, and the trend of the measurement. In the center of the interface is the central reactor panel. From top down, it contains the following components: 1) the annunciator panel showing the warnings present in the nuclear power plant; 2) the control rods that control the reaction within the reactor; 3) the safety injection trains; 4) the reactor gauges; and 5) the chatbox to communicate with the supervisor or respond to questions.

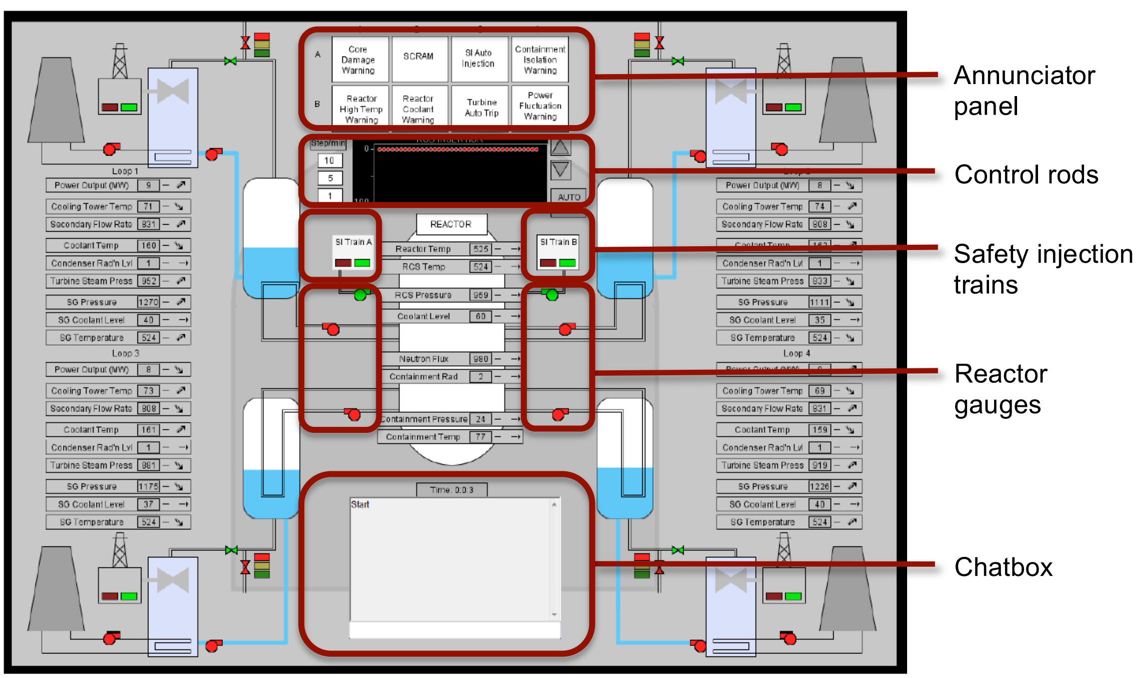


Figure 5-1: The HOMER primary interface

In the experiment, the primary task of the human operator was to monitor the primary control interface for an alarm condition. When an emergency event happens, the triggered alarm is shown on the annunciator panel. The operator needs to control the pumps, valves, and control rods following the procedure (referenced via the safety manual) to resolve the alarm. The safety manual was provided in a binder, which is typical of actual operations and this binder contained the procedures required for each type of alarm. Under normal conditions with no alarm, the operator needed to record specific parameters from the primary control interface to a secondary interface every 30 minutes, and respond to chat message queries spaced roughly every 20 minutes. Ratings of boredom, workload, and fatigue were each collected on a 5-point Likert scale at 30-minute intervals.

5.1.2 Experiment Design

The experiment had three independent variables: operating environment, critical event onset time, and secondary task availability. Operating environment had two levels: sterile and with unrestricted distractions. In the unrestricted distractions condition, participants were free to use cell phones, read books, browse the Internet, or do whatever else they liked, so long as they did not physically prevent other participants in the room from doing their tasks. In the restricted distractions (sterile) condition, participants were prohibited from bringing cell phones, books, electronic devices or any other type of distracting item into the experimental area. Internet access was also eliminated.

There were three levels of alarm onset time: early (alarm onset after 1.5 hours), middle (alarm onset after 2.5 hours), and late (alarm onset after 3.5 hours). These three alarm onset times were used in order to investigate whether a longer onset time of an alert influenced operator performance while responding to the emergency event.

The last independent variable had two levels: secondary task available and unavailable. The secondary task was a smaller nuclear power plant simulation on a separate display with one loop starting in a "cold shut-down" mode with no power being produced. Participants manipulated the controls in order to produce maximum power. This secondary interface represented a gaming interaction that gave operators opportunities to allow themselves to be distracted, but in an environment that was close to their operational one.

In addition to the independent variables, event scenario was also varied. Three different alarm scenarios were created based upon actual emergency procedures. Each was designed to take the same amount of time to complete, while requiring different paths to arrive at a successful solution. Each scenario required between 30 to 34 interactions with the interface to clear the alarm condition if the operating manual was followed correctly. Each participant cleared only one alarm, which was randomly selected out of the three scenarios. Participants were tested three at a time in the same room, each with a different event scenario. This made sure that they could not learn from each other to find the solution for clearing the alarm.

Dependent variables were measured in several categories, including primary task performance, attention states, subjective boredom, fatigue and workload ratings during the experiment, as well as self-assessed post hoc performance. The primary task performance was based on whether the alarm was cleared or not and the time required to clear the alarm. The attention states of participants were determined by coding the video recordings of individual participant behavior based on the definitions similar to the ones described in Chapter 4. The three main states were directed, divided and distracted:

1) Directed State: The participant faced the primary monitor, which contained the main control interface.
2) Distracted State: The participant did not face the computer monitor that contained the main control interface and attention was clearly directed elsewhere.
3) Divided State: The participant faced the main computer monitor but also engaged in another activity such as talking, eating, or drinking.

When the attention was directed at the secondary screen, it was considered a subcategory of distracted attention, called secondary distracted state. Primary task performance and attention states were used as main metrics for the comparison between experimental data and system dynamics model output.

Personality factors were measured with the Boredom Proneness Scale and the NEO-FFI. Participants also rated their sleep for the past two nights, overall health, game experience, computer skills, and perception of nuclear power plants.

5.1.3 Results

The success rate in clearing the alarm is presented in Table 5-1. The average alarm clearance time in each condition is shown in Figure 5-2 along with the standard deviation. In the sterile operating condition, 14 participants (77%) were able to clear the alarm as compared to

8 participants (44%) in the distraction condition. The difference between completion rates is significant ($\chi^2 = 4.0$, df = 1, p = 0.046), calculated using a logistic regression model. The average alarm clearance time was consistently higher when distractions were available, although an ANOVA test shows that the difference is not statistically significant.

Table 5-1: Success Rate in Clearing the Alarm

	Early (1:30 Event Onset)	Middle (2:30 Event Onset)	Late (3:30 Event Onset)
Sterile	100% (6/6)	83% (5/6)	50% (3/6)
Distractions	83% (5/6)	33% (2/6)	17% (1/6)

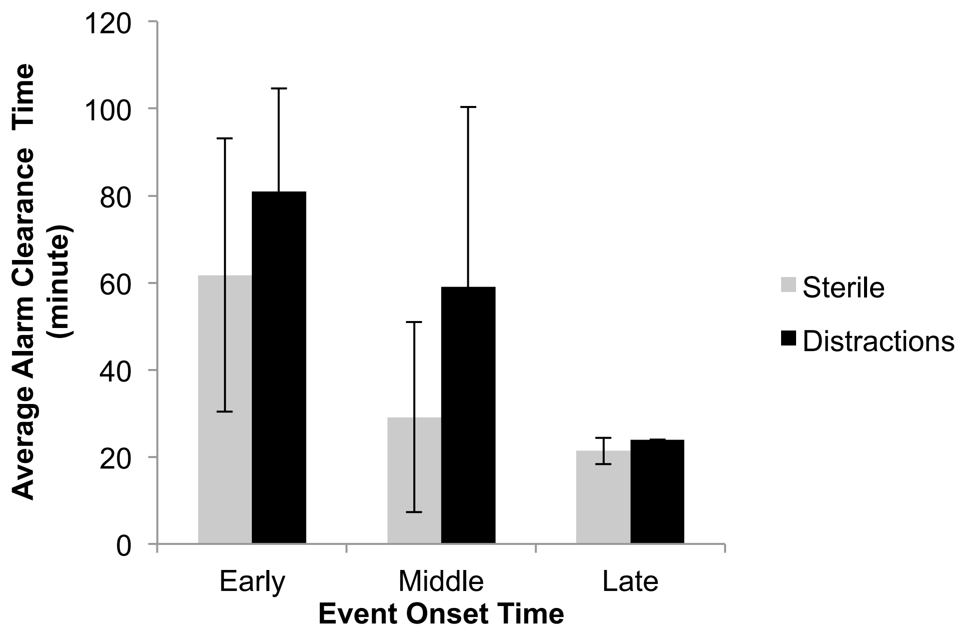


Figure 5-2: Average Alarm Clearance Time

Event onset time also affected the success rate in clearing the alarm. 11 participants (91.7%) in the early onset group cleared the alarm while only 7 (58.3%) cleared the middle onset group, and 4 (33.3%) in the late onset group. The difference between completion rates is significant ($\chi^2 = 6.6$, df = 2, p = 0.037), calculated using a logistic regression model. Average alarm clearance time decreased with increased alarm onset times. The ANOVA test show that the difference is significant (F(2,16) = 4.96, p = 0.021). This decreasing trend may be due to several reasons. First, the late onset group was only given 30 minutes to clear the alarm while the actual clearance time could have taken much longer than that. Second, there could be a learning effect. As participants spent more time monitoring the system before the alarm onset, they may get more familiar with the system, interface and the safety manual, causing a decrease in alarm clearance time. Third, participants with earlier onset times may perceive a lower level of stress because they felt there was more time to solve the problem.

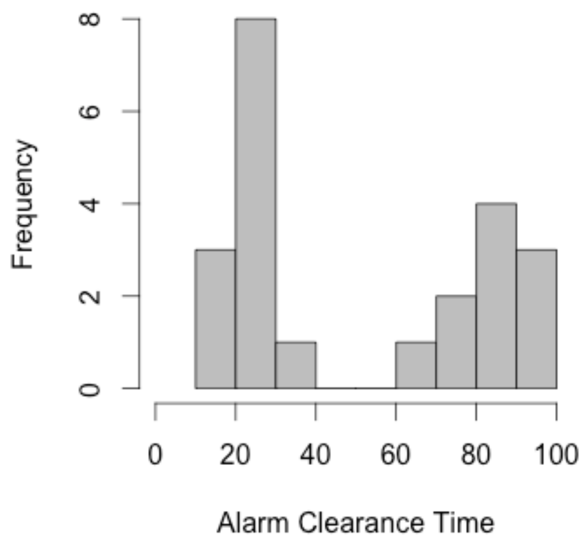


Figure 5-3: Histogram of Alarm Clearance Time

In addition, the alarm clearance time also had large standard deviations within each condition, in part due to the small sample size of the experiment. A further investigation of the clearance time shows that it is almost a bimodal distribution, as shown in the histogram in Figure 5-3. One reason is the familiarity with the system. All the participants in the experiment were students. Although they went through training before the test session, their familiarity with the system was low compared to real operators, which would lead to large standard deviations. In the post experiment survey, some participants commented that they could not follow the procedures in the binder. Clearing the alarm requires comprehension of the procedures, interface, and the system. If a person did not understand the system, he or she might take a long time to locate the correct procedure to follow. In addition, an error made in an earlier step in following the procedure may cause delays for the operator, or result in an incorrect attempt to solve the problem. This could result in significant increase in alarm clearance time as the operator may need to rely on his or her own understanding to get back on track. The bimodal distribution of alarm clearance time could be due to individual differences in problem solving skills. However, this hypothesis needs further validation.

Problem solving skills of the participants were estimated based on their experience in playing "problem solving" games (strategy, puzzle, real time strategy games) as reflected in pre-experiment survey. Participants rated the frequency of playing games on a 4-point Likert scale (*Game Frequency*), and listed the type of games they often played. The games they played were classified into two categories and represented by a binary variable, *Problem Solving Game Indicator*. *Problem Solving Game Indicator* equaled one if the game was a strategy, puzzle, or real time strategy game. These are the games that require problem-solving skills. Otherwise, *Problem Solving Game Indicator* equaled zero. Experience in playing "problem solving" games (*Game Experience*) was calculated as:

$$Game\ Experience = Problem\ Solving\ Game\ Indicator \times Game\ Frequency \tag{36}$$

In other words, *Game Experience* was determined by whether they play "problem solving" games and how often they play games. The impact of such game experience on performance was tested using a multinomial logistic regression.

The dependent variable was *Clear Class*, which has three levels:

- Fast: *Clear Class* = 0, the alarm was cleared within 30 minutes.
- Slow: *Clear Class* = 1, the alarm was cleared using more than 30 minutes.
- Failed: *Clear Class* = 2, the alarm was not cleared.

Participants who did not clear the alarm in the late event onset condition were removed. Since they had only 30 minutes from the event onset to the end of the experiment to solve the problem, it was unclear whether they actually belong to the Slow or Failed class. The reference level in the logistic regression was set to Slow (*Clear Class* = 1) to estimate the odds ratio. The independent variables were Game Experience (ranging from 0 to 5), Event Onset Time (90, 150, 210 minutes), and Distraction Condition (Sterile, Distractions). Results of the multinomial logistic regression are presented in Table 5-2.

Comparing the Fast class to the reference level of Slow class, Game Experience was significant ($p = 0.041$). If a participant had a *Game Experience* value higher by one unit, the odds of being in the Fast class rather than the Slow class would be 11.295 times higher. This means that for those who actually cleared the alarm, the more often they played "problem solving" games (strategy, puzzle, real time strategy), the faster they cleared the alarm. Game Experience was not significant when comparing the Failed class to Slow class. These analyses shows that experience in playing problem solving games could affect the time a person need to clear the alarm.

Table 5-2: Multinomial Logistic Regression for Clear Class

		B	Exp(B)	Std. Error	z	p
Fast (Clear Class 0)	(Intercept)	-11.759	0.000	4.650	-2.529	0.011*
	Game Experience	2.424	11.295	1.186	2.044	0.041*
	Event Onset Time	0.081	1.084	0.031	2.636	0.008**
	Condition	-0.712	0.491	1.909	-0.373	0.709
Failed (Clear Class 2)	(Intercept)	-9.563	0.000	3.889	-2.459	0.014*
	Game Experience	1.812	6.125	1.236	1.466	0.143
	Event Onset Time	0.055	1.057	0.025	2.195	0.028*
	Condition	2.748	15.614	1.837	1.496	0.135

Attention was another important dependent variable. As stated earlier, attention states were classified into three categories based on video recording. Overall, the participants spent 49% of their time in directed attention state, 7% in divided attention state, and 44% in distracted attention state. In order to clear the alarm, participants must spend time looking at the procedures, resulting in an increase of divided attention after the event onset. The attention state distribution before the event onset shows a slightly higher percentage of directed attention (53.1%) and slightly lower divided attention (3.7%) as compared to the overall distribution.

When comparing the attention distribution before event onset in the two operating conditions (sterile or with distractions), only the divided attention state differed significantly ($F(1,31) = 8.47$, $p = 0.007$). It seems that the operators were willing to split their attention between the primary task and other activities, rather than being fully distracted. When comparing the attention distribution before event onset in the three event onset groups, only directed attention significantly differ across the groups ($F(2,31) = 5.55$, $p = 0.009$). The percentage of distracted attention is marginally significant different across the three event onset groups ($F(2,31) = 3.1028$, $p = 0.059$). The average percentage of each attention state is listed in Table 5-3. The data shows a lower level of directed attention in early event onset group. While an average percentage in each attention state was used for statistical analysis, detailed time series data was used in the system dynamics model. The level of attention in these three states averaged over 15-min intervals is presented in Figure 5-4.

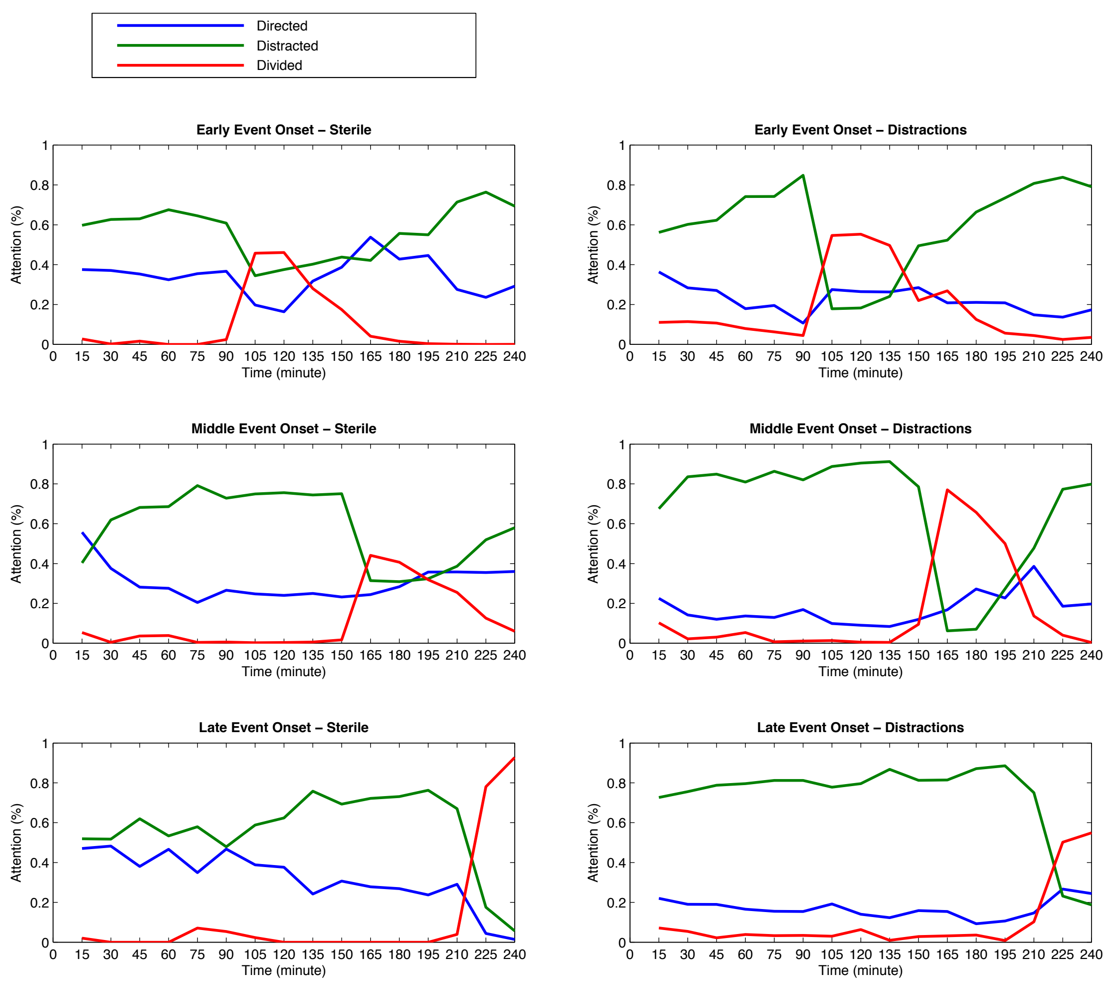


Figure 5-4: Attention States under Each Condition

Secondary task availability did not influence the primary task performance on either success rate or alarm clearance time. It did not influence the distribution of attention states significantly either. In later analysis, secondary task availability was not used as an independent variable.

Table 5-3: Distribution of Attention States before Event Onset

Condition	Time	Directed	Divided	Distracted
Sterile	Early	45.3%	0.2%	54.2%
With Distractions	Early	43.6%	16.0%	40.3%
Sterile	Middle	58.1%	2.9%	38.7%
With Distractions	Middle	59.6%	5.2%	34.8%
Sterile	Late	53.2%	0.3%	45.9%
With Distractions	Late	51.3%	3.6%	44.9%

Boredom, fatigue and workload were rated every 30 minutes during the experiment. The repeated boredom ratings were significantly higher in sterile condition comparing to distraction condition ($F(1,29) = 9.91$, $p = 0.004$). Similarly, repeated fatigue ratings were significantly higher in sterile condition as well ($F(1,29) = 9.63$, $p = 0.004$). Fatigue ratings also showed a marginal significant difference across the three event onset groups ($F(2,29) = 2.76$, $p = 0.080$), with the middle event onset time having the highest fatigue rating. Workload rating did not differ across groups.

5.2 Model Parameters

In order to test whether the system dynamics model in Chapter 3 could capture the change of human attention and performance in responding to an emergency event, the model outputs were compared with the experiment data. The parameters used in the model are shown in Table 5-4. *Baseline Self-Imposed Event Rate* was set to zero in this case, as extra interactions with the system are negligible. Under normal condition, the operators could change the controls and settings of the nuclear power plant to vary the power output. However, these parameters were already managed and optimized by the automation. Human operators were not motivated to change these unless it was ordered by their supervisor.

Exogenous Event Rate is based on the event onset time in the experiment. The PULSE function in Vensim® is used to model the arrival of one emergency event at the onset time at 90 minutes, 150 minutes or 210 minutes from the start of the experiment. *Normal Processing Rate* equals 0.05 tasks/minute, because each emergency event was designed to take about 20 minutes to process. In the model, event processing was stopped after the *Required Processing Time* was all used. In this task scenario, *Required Processing Time* was set for 98.23 minutes, which is the maximum alarm clearance time in the experiment data. Although the emergency event was designed to be cleared in about 20 minutes, it took significantly longer for the experiment participants to clear the alarm. Since the participants were allowed to continue handling on the emergency event even if they went beyond the 20 minutes, the maximum clearance time was used as an estimation of *Required Processing Time* instead of 20 minutes. Processing time is individual dependent.

Using a single constant value is not sufficient to capture individual differences or the range of possible behavioral outcomes. This is a limitation of system dynamics modeling in general, which is addressed in more detail in the last chapter. *Average Processing Rate* was calculated as an exponential moving average of *Event Processing Rate*. To reflect potential changes in task load, the sampling time interval for the moving average (*Time Window*) was set at 5 minutes.

For *Executive Control Resource*, both the *Depletion Time* and *Restore Time* are set to 240 minutes, which is the length of the mission in the experiment. For attention, Attention on Task is initialized at 0.4, because the percentage of time in the state of directed attention is about 40% at the beginning of the experiment. The initial value and the maximum level of *Executive Control Resource* are both set to 0.4 to be consistent with the initial attention level. The minimum level of *Executive Control Resource* equals zero. *Average Time to Distract* is 25 minutes, same as in Chapter 4, because previous research shows that sustained attention decreases around 20-30 minutes into the mission (Mackworth 1957). To incorporate the impact of external distraction sources, *Time to Distract* is calculated as in Equation (37). *Sources of Distraction* describes the level of external distractions ranging from zero to one. For the operating condition with distractions, *Sources of Distraction* was set to 0.9. For the sterile condition, *Sources of Distraction* was set to 0.1. With more sources of distraction, it takes a shorter time for people to get distracted.

$$Time\ to\ Distract = Average\ Time\ to\ Distract * (1 - Sources\ of\ Distraction) \tag{37}$$

In the scenario described previously, divided attention between the procedures and the control interface was needed in order to process the emergency event. As a result, divided attention increased rapidly after the event onset as shown in Figure 5-5. Before event onset, divided attention was at a low level close to zero. After event onset, divided attention rapidly increases to a high level around 0.6. The starting times of such increases and the times reaching the first local maximum points are marked on Figure 5-5. This data can be used to estimate the value of Time to Refocus in the model. Divided attention took about 7 minutes on average to increase from the low level to high level for the early event onset group, 5 minutes for the middle event onset group, and 6 minutes for the late event onset group. Since these numbers were around the same range and the model needs to fit the experiment data across all conditions, Time to Refocus was set to 6 minutes in the model by averaging the three numbers.

Table 5-4: Model Parameters

Model Parameters	Values	Parameters for Nonlinear Relationships
Baseline Self-Imposed Event Rate	0 tasks/minute	$k_1 = 1.55$
Exogenous Event Rate	PULSE(Pulse Time, 1) Pulse Time = 90, 150 or 210	$k_2 = 6.11$ $c_1 = 1$
Normal Processing Rate	0.05 tasks/minute	$c_2 = 2$
Required Processing Time	98.23 min	$k_3 = 10$
Time Window	5 min	$k_4 = 1$
Depletion Time	240 min	$k_5 = 4.13$
Restore Time	240 min	$c_5 = 1.91$
Executive Control Resource: Initial value Maximum Level Minimum Level	 0.4 0.4 0	$m_5 = 0.08$
Attention on Task, Initial value	0.4	$k_6 = 4.90$
Average Distraction Time	25 min	$c_6 = 9.17$
Refocus Time	6 min	$k_7 = 1.72$
Source of Distraction	0.9 for with distractions	$c_7 = 4.76$

There are other parameters used in the model to describe the nonlinear relationships as presented in Chapter 3. The values of these parameters were chosen using the model calibration function in Vensim® based on the data in the distraction condition. The values of these variables are in the last column in Table 5-4. These nonlinear relationships are visualized in Appendix A. Data in the sterile condition was not used for model calibration. They were saved for testing the model's prediction ability when the system was changed (restricting external distractions). From the modeling perspective, this ensures the generalizability of the model and avoids overfitting to only one data set. From the design perspective, designers usually do not know how the system will behave with the proposed system changes. They need to rely on existing data to estimate the impact of new system changes. Calibrating the model using data under the distraction condition and predicting the outputs when distractions were restricted follows the same process as the designers.

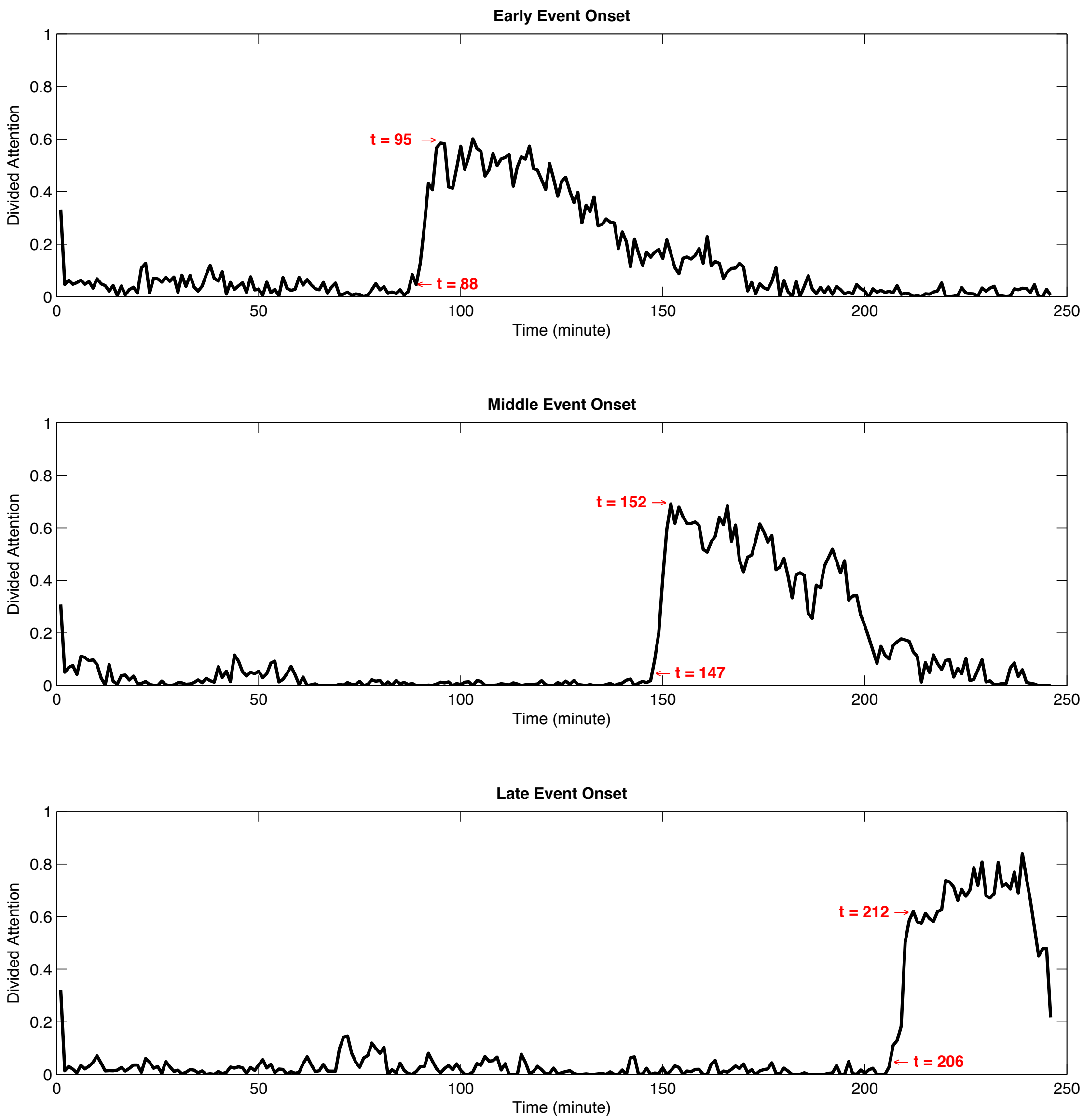


Figure 5-5: Divided Attention with Different Event Onset Times

5.3 Model Fit

In order to test whether the model could successfully capture human performance and attention in responding to an emergency event after a long period of relative inactivity, two key outputs of the model are compared with the experiment data. A quantitative assessment of the model's fit to experimental data under the condition with distractions is provided in Table 5-5.

Table 5-5: Simulation to Experimental Data Fit Statistics (Distraction)

Summary Statistics	Attention on Task (Distractions)		
	Early Onset	Middle Onset	Late Onset
Coefficient of Determination (R^2)	0.540	0.049	0.118
Root Mean Square Error (RMSE)	0.160	0.181	0.128
Mean Square Error (MSE)	0.026	0.033	0.016
Bias component of MSE (U^M)	0.003	0.107	0.560
Variation component of MSE (U^S)	0.056	0.021	0.037
Covariation component of MSE (U^C)	0.942	0.872	0.403

Attention on Task is compared with the sum of directed and divided attention in the experiment data. Divided attention was included as a part of attention on primary task, because participants switched between the computer screen and the safety procedures in a binder when they were trying to clear the alarm. Percentage of time in the state of directed and divided attention in the experiment is summarized by minute, as shown in the blue line in Figure 5-6. The red line shows the change of *Attention on Task* in the system dynamics model. In general, the model could capture the decrease of attention under normal conditions and the increase of attention with emergency event onset.

The best fit was the early onset time group, with a R^2 value of 0.540. R^2 values for middle and late onset groups are low, likely influenced by the large variation of attention, which is not captured in the model. The largest component of MSE is U^C, which equals to 0.942, also indicating a good fit. For the middle onset group, U^C is 0.872, which is still the largest component of MSE. This means that most of the error was due to unsystematic, or random, variation, indicating a good model fit. For late onset, U^M equals to 0.560, which is the largest component of MSE. This indicates a bias from the experiment data. However, visually inspecting the two lines in Figure 5-6 does not reveal a large difference.

The overall fit of the model is affected by several factors. First, the sample size is small for each condition. The attention data is averaged across 5 or 6 people in each condition. Second, human behavior usually has large variation. The model fit could potentially be improved with a larger sample and recalibration. Since the purpose of the model is to capture the general trend of attention change rather than to predict point-to-point attention state, the current model fit is good enough for this purpose.

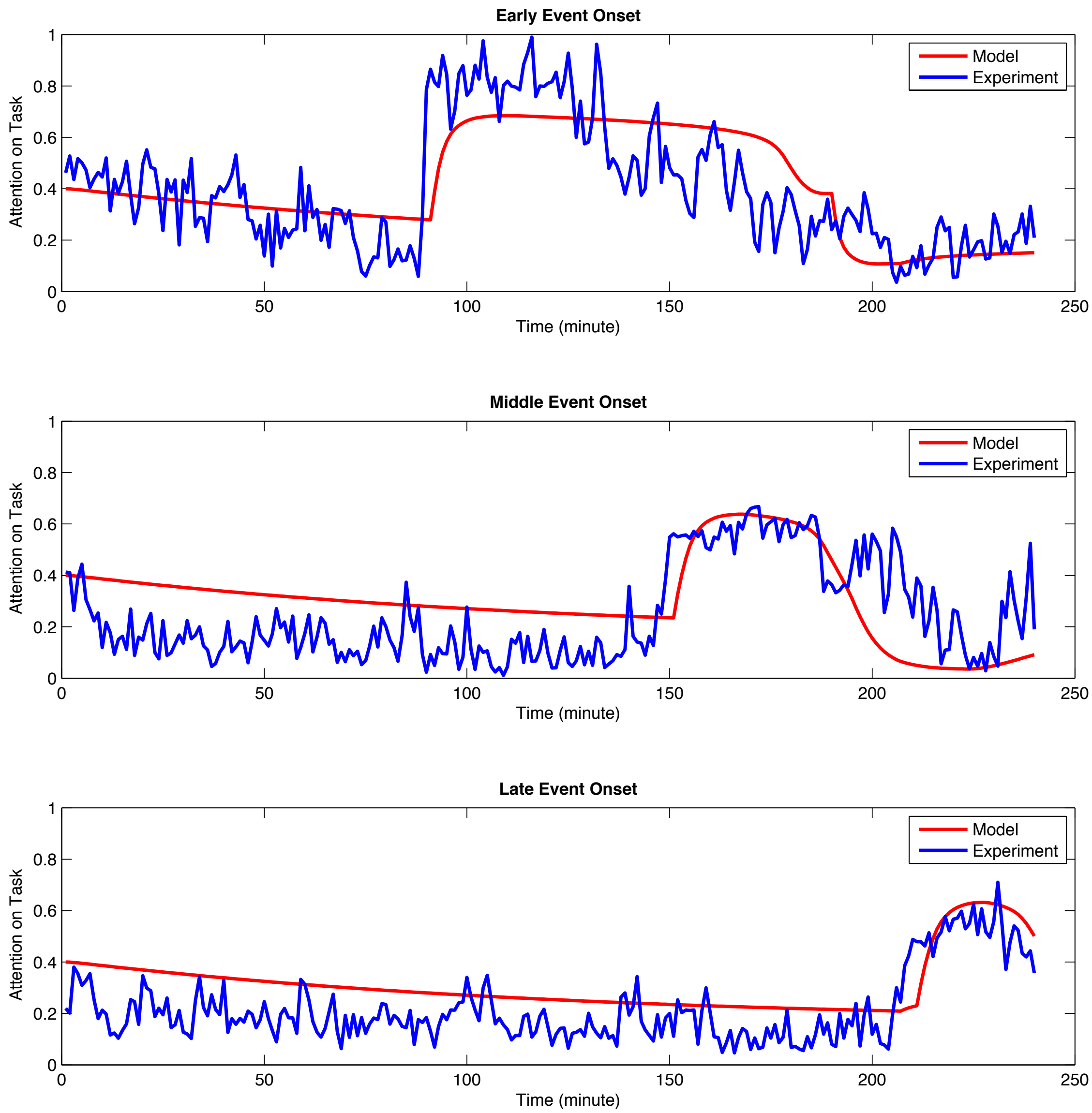


Figure 5-6: Comparison of Attention on Task under Distraction Condition

Dynamic hypothesis 2, says the reduction of executive control resource and attention on a primary task under low task load leads to worse human performance in dealing with unexpected tasks. In order to test this hypothesis, the model fit on performance is also evaluated, as shown in Figure 5-7. Alarm clearance success rate in the experiment is compared with *Events Processed* in the model. Overall, the model provides a good fit to performance in terms of the alarm clearance success rate. In both the experiment and the simulation model, the success rate is lower with later event onset times. This is due to the lower level of attention and executive control resource. In other words, hypothesis 2 is supported.

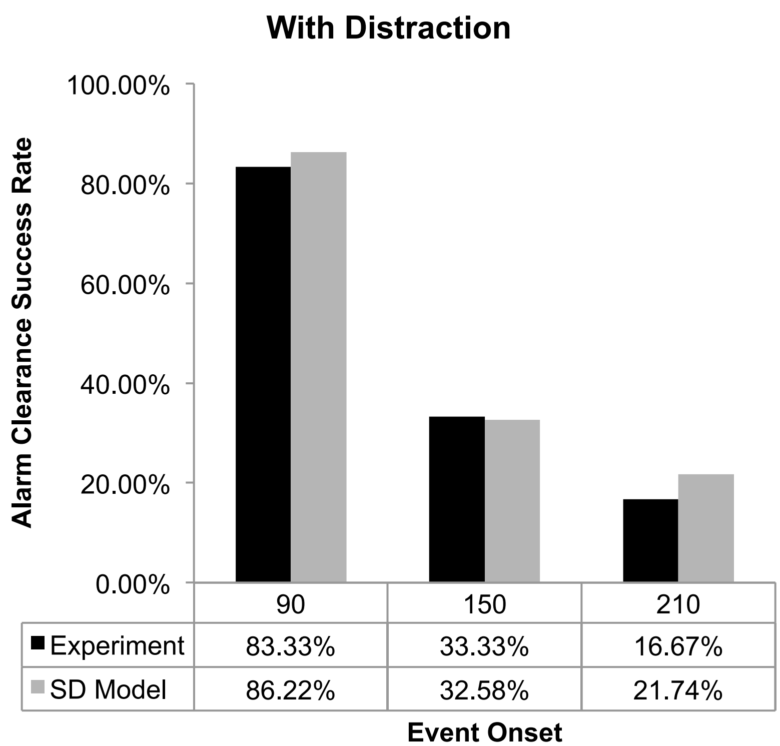


Figure 5-7: Comparison of Alarm Clearance Success Rate under Distraction Condition

5.4 Predictive Validation: Effect of Restricting External Distraction

In order to evaluate the model's ability in predicting system design changes, the effect of restricting external distraction sources was predicted using the model. All the parameters calibrated using data under the distraction condition were kept the same. Two parameters were changed to reflect the impact of restricting external distractions sources. *Source of*

Distraction was decreased from 0.9 to 0.1. The initial value of *Executive Control Resources* and *Attention on Task* were slightly increased from 0.4 to 0.45 since it was assumed that people would pay more attention to the task if external distraction sources were eliminated.

Table 5-6: Simulation to Experimental Data Fit Statistics (Sterile)

Summary Statistics	Attention on Task (Sterile Condition)		
	Early Onset	Middle Onset	Late Onset
Coefficient of Determination (R^2)	-0.105	-0.227	0.045
Root Mean Square Error (RMSE)	0.162	0.163	0.119
Mean Square Error (MSE)	0.026	0.027	0.014
Bias component of MSE (U^M)	0.391	0.387	0.006
Variation component of MSE (U^S)	0.000	0.002	0.002
Covariation component of MSE (U^C)	0.609	0.611	0.993

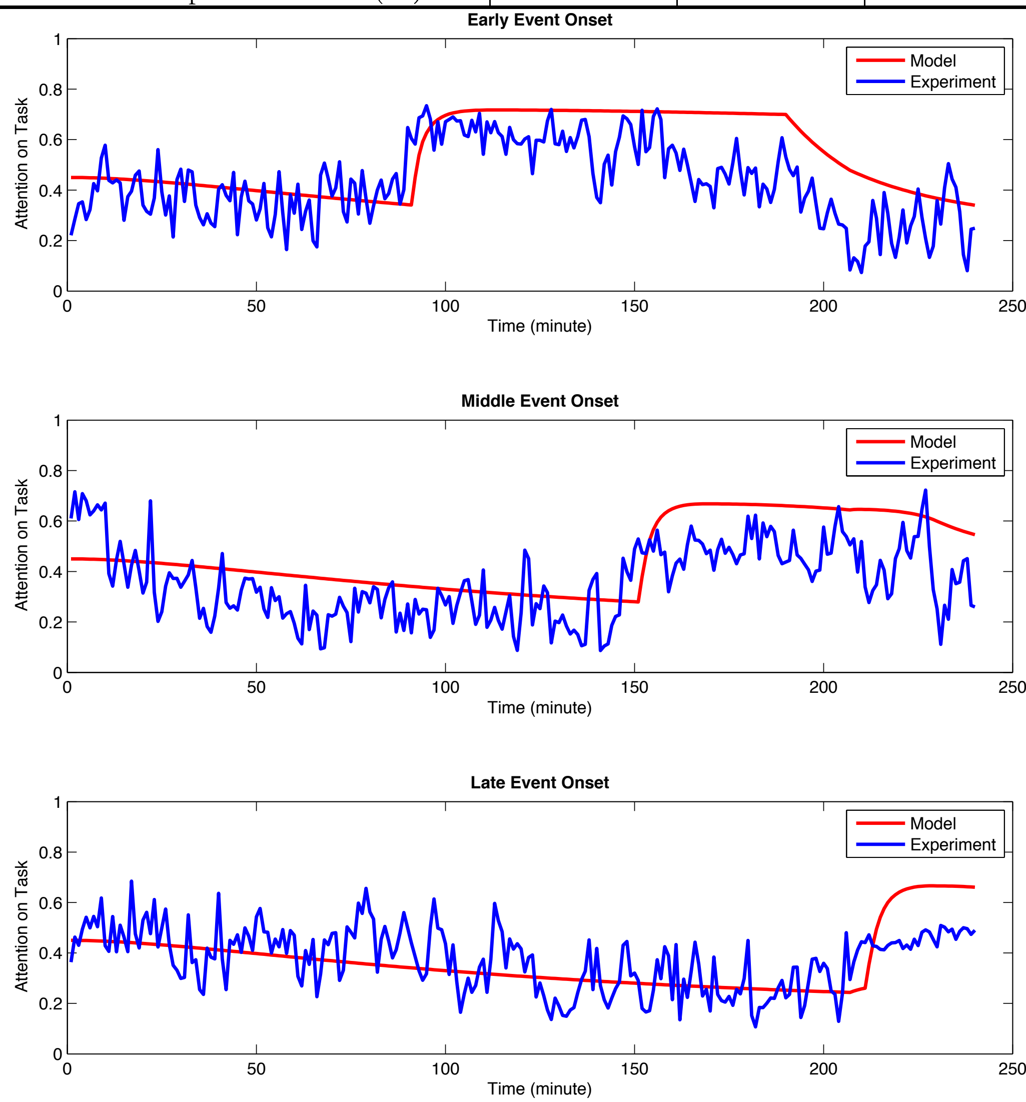


Figure 5-8: Comparison of Attention on Task under the Sterile Condition

The model outputs with the above changes were compared with experiment data under the sterile condition to see whether the model predictions were reasonable. The quantitative statistics of model fit are presented in Table 5-6. Overall, R^2 values are very low. However, examining the values of U^C shows that it was the largest component of MSE under all three conditions, which indicates a good model fit.

Visually inspecting the model fit as shown in Figure 5-8, the model is able to capture the major changes of attention. The model overestimates the level of attention at the end of the experiment in the late onset condition. As with the previous results, the overall model fit for attention is affected by the small sample size, variation of human behavior, and the exclusion of small variations of attention in the model.

The performance prediction in terms of alarm clearance success rate was also evaluated, as shown in Figure 5-9. The model slightly underestimates the performance in the middle onset condition. The biggest difference between the model output and experiment data is for the late onset condition with an alarm clearance success rate of 50% for the experiment data, while the model predicted 26.99%.

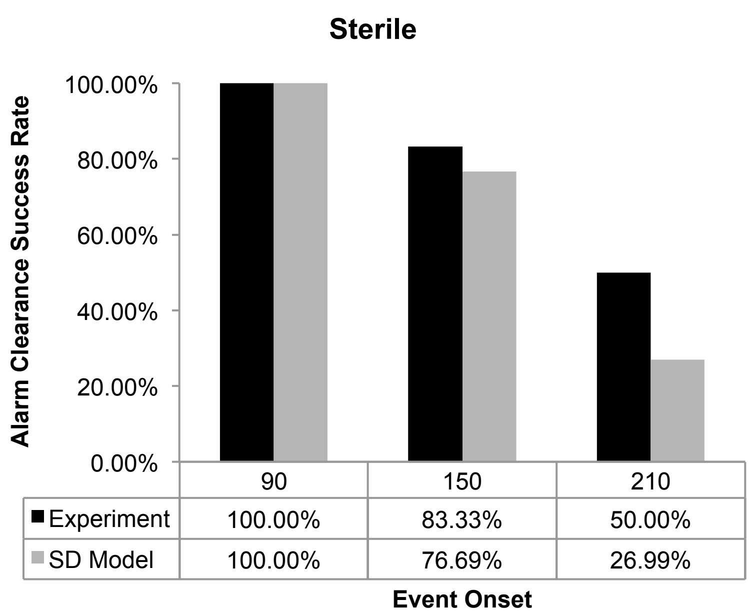


Figure 5-9: Comparison of Alarm Clearance Success Rate under Sterile Condition

In order to investigate the large difference in performance in the late onset condition, the alarm clearance time was investigated. As stated earlier in section 5.1.3, alarm clearance time seems to have a bimodal distribution, with one group of people clearing the alarm in about 22 minutes and the other group in about 80 minutes. The participants who cleared the alarms under each condition were classified into two classes based on their alarm clearance time (fast and slow), as shown in Figure 5-10. There were a total of six participants under each condition. When distractions were available, only two people cleared the alarm in the fast class, with one in the middle onset condition and one in the late onset condition. When the distraction sources were restricted, more people were able to clear the alarm fast. Two out of six participants cleared the alarm in the early onset condition, four in the middle onset condition, and three in the late onset condition. Further, for the late event onset group, only half an hour was allowed to clear the alarm before the experiment ended. This means only people who could clear the alarm fast could finish the task, while those in the slow class (Class 2) would not be able to clear the alarm. This is also reflected in the data as shown in Figure 5-10. All the participants who cleared the alarm in the late onset condition were in the fast class (Class 1).

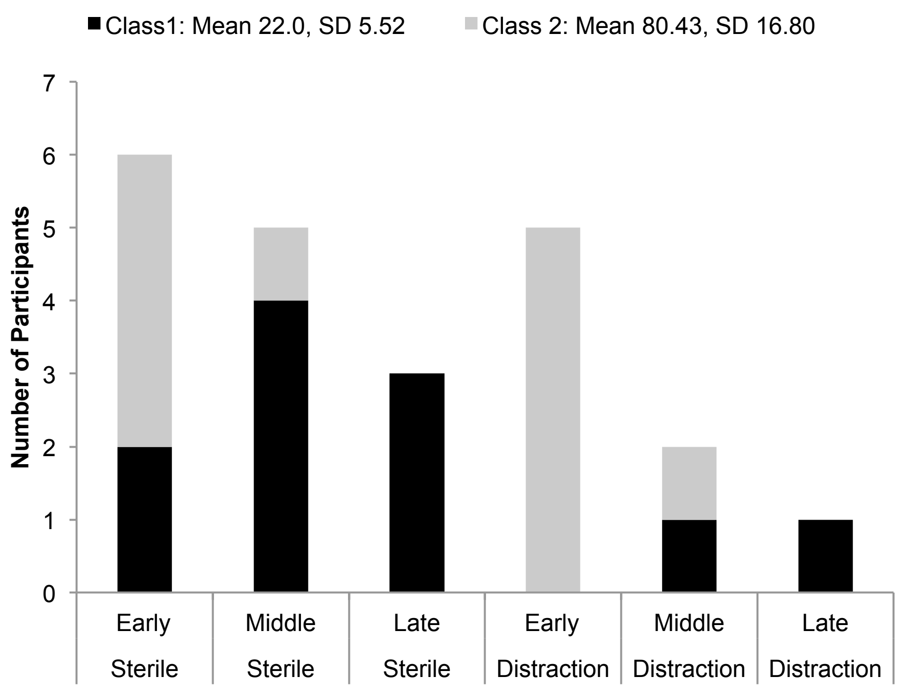


Figure 5-10: Alarm Clearance Class for Each Condition

The bimodal distribution of alarm clearance time is likely to be the reason for the under-prediction of performance in the sterile, late onset condition. In the model, Normal Processing Rate was set to be a single constant number. In the system dynamics model, this corresponds to a homogeneous group of individuals whose task processing is all around this average number. This is not consistent with the bimodal distribution as reflected in the experiment data, and is illustrated in Figure 5-11. The x-axis represents time, and the y-axis represents the overall task processing progress of all the participants. The slope represents the rate of task processing. Instead of following a straight line of task processing as in the model, it should be a broken line with two different slopes representing the two classes of alarm clearance time in the averaged experiment data. Those who clear the alarm fast caused a steep increase towards task completion at the beginning. After they completed the task, the slower people continued to work on clearing the alarm following the flatter line. With late event onset, the experiment ended 30 minutes after event onset. This allowed the fast ones to clear the alarm as they took about 20 minutes. However, the slower ones could not finish clearing the alarm before the experiment ended. As illustrated, the dotted line for experiment ending time was around the intercept of the two line segments. The bimodal distribution and the experiment end time together cause the gap between experiment data and model output. This also reflects the limitation of the system dynamics model in capturing the heterogeneity in population.

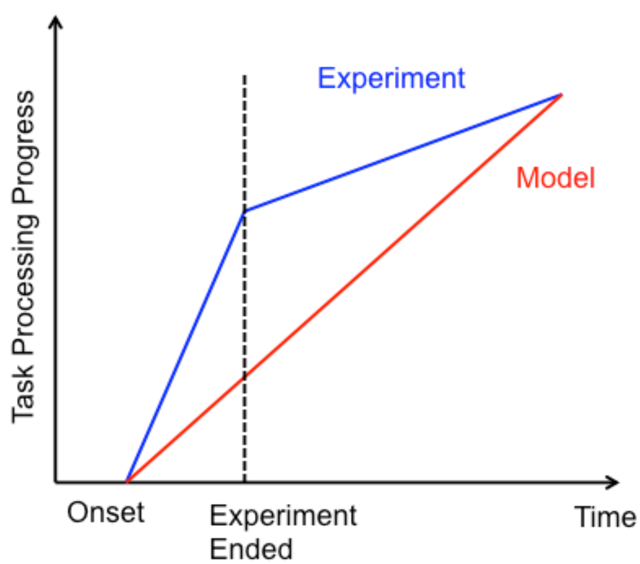


Figure 5-11: The Impact of Bimodal Clearance Time on Model Fit

In fact, by separating the participants into two homogenous groups, setting different processing rates for the two groups and different percentages of people in the two groups for each condition, the model could provide a better fit with the experiment data after recalibration. However, since the goal is to build a generalized model for attention and performance in low task load environments, we kept the model simple. Although the analysis shows that gaming experience in problem-solving games influenced the alarm clearance time, the reason for the bimodal distribution still needs further investigation. There might be other reasons in addition to individual differences that cause such bimodal distribution. What's more, even if individual differences are the solely reason, the percentage of people in each group (fast or slow) is likely to change in different environments. The bimodal distribution is likely related to the fact that participants were mostly students with limited training (novice). With experienced professional operators, variance on clearance time would be much smaller. If the participants were all experts familiar with the safety procedures, the processing time would be consistently low. The PAL model could be applied to such a case by setting a short constant processing time.

Overall, the model predicted that performance would be better if external distraction sources were restricted. The performance still decreased with later event onset times. These two conclusions were consistent with the experiment data. The underestimate in late onset condition maybe due to the fact that alarm clearance time was bimodally distributed within the experiment participants, which may be an experiment artifact.

5.5 System Improvement Prediction: Testing Task

One purpose of building the PAL model is to use it to facilitate the system design process. Instead of running experiments with humans to test each system design alternative, the PAL model can be used to identify designs that could lead to improvements on performance and attention. Although additional studies are needed to evaluate the designs suggested by the model, development time and cost can still be greatly reduced. This section demonstrates evaluating the impact of an additional testing task in a nuclear power plant using the PAL model.

5.5.1 Testing Task Description

Nuclear power plant monitoring is a complex task. It requires familiarity with the system and safety procedures, high situation awareness, good problem solving skills, and vigilance to detect and resolve the alarms (Mumaw et al. 2000). In the experiment described in this chapter, attention on the primary task decreased over time due to the low task load from passive monitoring. This result was replicated in the PAL model.

One method that might improve attention management and performance in alarm clearance is to increase task engagement by adding system-testing tasks under normal operating condition. For example, the operator can manipulate the plant system or sub-systems to see whether the indicators respond as expected. This has two potential benefits. First, the task load during the monitoring period can be increased, which may reduce boredom and lead to better attention management. Second, actively testing system components could increase familiarity with the system and situation awareness of the current state. This may result in faster detection of a problem, faster diagnosis of the cause of an alarm, or better compliance with the safety procedure. For example, if a certain indicator did not respond as expected during the testing procedure, this could reveal potential safety problems sooner.

5.5.2 The Impact of Testing Task

The effect of this approach on attention management and performance was tested using the PAL model. In the model, a system-testing task was introduced at 30 minutes from the start of the mission as a single pulse. It was assumed that the testing task was easier as compared to clearing the alarm. As a result, the pulse size of the testing task was set to 0.25, as compared to 1 for the alarm clearance task, but could take other values based on the difficulty of the testing task. The testing task was required to be completed within 30 minutes. The original model outputs that replicated the experiment data were used as baseline. Two scenarios of adding a testing task were evaluated using the model. In the first scenario, a testing task was included without any other changes. In the second scenario, the testing task resulted in a 30% reduction in alarm clearance time, due to better familiarity with the system, better situation awareness, and faster diagnosis of the root cause of the alarm. The impact of testing task on performance is summarized in Table 5-7, and the impact on attention is presented in Figure 5-12.

Table 5-7: Impact of Testing Task on Performance

Condition		Alarm Clearance Rate		
		Baseline	Testing Task	Testing Task with Performance Gain
Sterile	Early	1.000	1.000	1.000
	Middle	0.767	0.903	1.000
	Late	0.270	0.273	0.398
Distractions	Early	0.862	1.000	1.000
	Middle	0.326	0.504	0.693
	Late	0.217	0.262	0.379

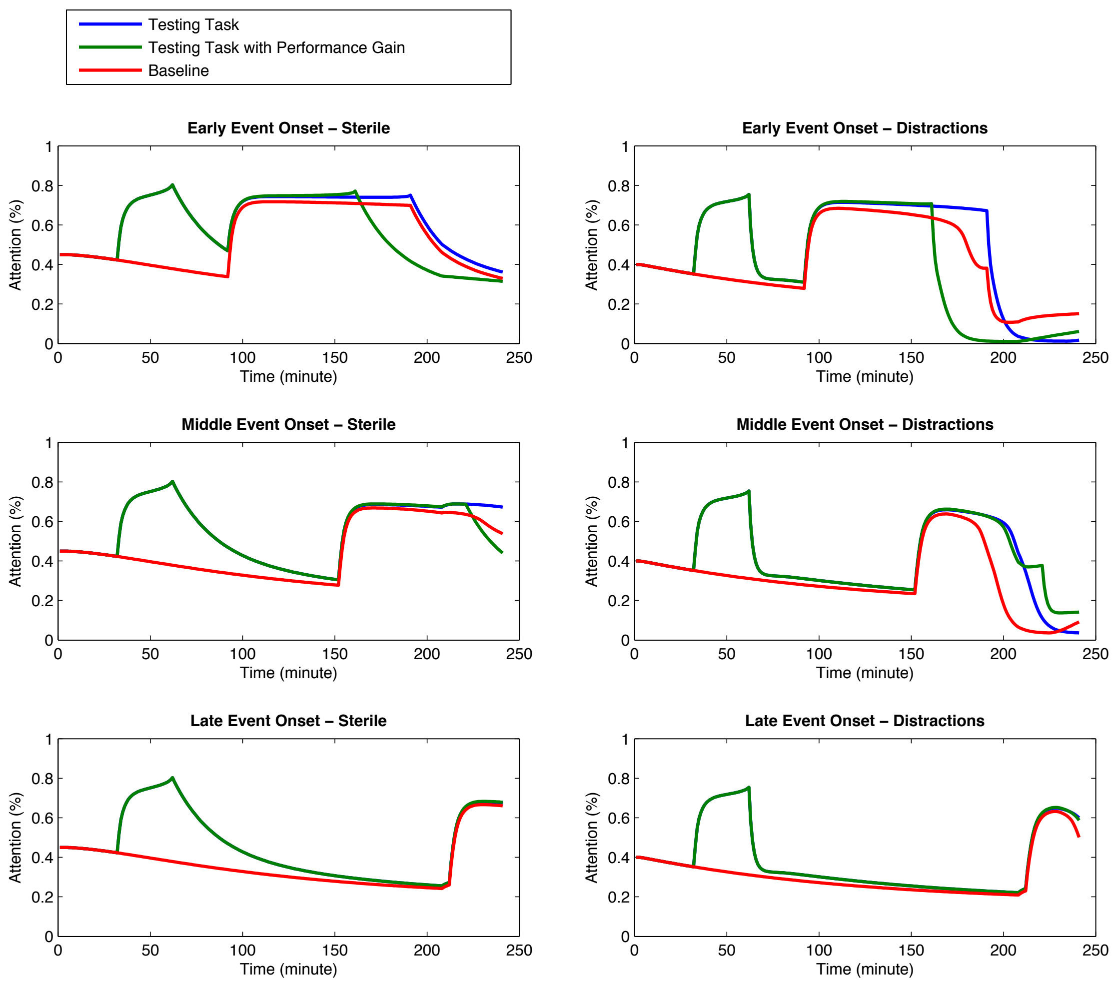


Figure 5-12: Impact of Testing Task on Attention

It can be seen that attention increased during the processing of the testing task and decreased afterwards. This results in a slight increase of attention level prior to alarm onset and predicted higher alarm clearance rates under all conditions. If alarm clearance time was shortened due to the testing task, performance was further improved, as listed in the last column in Table 5-7. What's more, the comparison among these three cases provides a breakdown for sources of improvements. For example, in the sterile-middle onset condition, the original clearance rate was 0.767. The clearance rate improved due to better attention management and slower depletion of *ECR* by 0.903 – 0.767 = 0.136. Improvement from the faster clearance time was 1 - 0.903 = 0.097. Analyzing the three onset conditions using this approach, it can be seen that the performance improvement for early onset conditions was small because the original performance was already high. For the middle onset conditions, improvements from both sources were of similar magnitude. For the late onset conditions, faster clearance time resulted in larger improvements on performance compared to attention management. This is because the effect of the testing task on attention already decayed by the time the alarm happened.

In summary, adding an additional testing task could help attention management and improve performance in clearing alarms as demonstrated using the PAL model. System designers could use the PAL model to explore wider design space or filter design options to find those with larger improvements. The model could provide suggestions for strategic improvements. However, the detailed task or interface design cannot be directly derived from the model.

5.6 Chapter Summary

In this chapter, a long duration human subject experiment with a shift from low task load to high task load was introduced. In this experiment, participants monitored the control interface of a nuclear power plant and responded to an emergency event by following the safety procedures in a four-hour mission. Three levels of event onset time were tested to investigate whether the duration of monitoring time influenced operator performance while responding to the emergency event. The system required only infrequent human interactions under normal conditions, but relatively complex problem solving when an emergency event happened. Revisiting the task categories discussed in Chapter 3, the emergency event belongs to the category with low frequency and high level of human effort in Table 3-1. Two phenomena were observed in the experiment. First, performance in handling the emergency event decreased when people experienced low task load for longer periods of time. Second, performance was improved if external distraction sources were restricted.

In order to assess the ability of the model to replicate these behaviors, model parameters were set based on experiment data, previous literature, and model calibration using part of the experiment data (data under distractions available condition). The outputs of the model on attention and performance were compared with the experiment data. The comparison shows that the model successfully replicated the experiment behaviors in attention change and decrease of performance with later event onset times. As a result, dynamic hypothesis two was supported, which stated that the reduction of executive control resource and attention on a primary task under low task load leads to worse human performance in dealing with unexpected tasks.

The model was further tested on its ability to predict the effect of system interventions. Experiment data under the sterile condition were used for this purpose. In this condition, sources of external distraction such as Internet, cell phones, and magazines were restricted. In the model, the parameters were unchanged except for slight changes to account for the difference in initial level of attention, executive control resources, and whether external distractions were restricted. Comparison between the model outputs and experiment data shows that the model provided a good prediction on performance for early and middle onset times, but poor fit for late onset times. Further investigation of experiment data shows that this may be the result of the bimodal distribution of alarm clearance time. However, the model predicted that the performance was improved when distraction sources were restricted, which is consistent with the experiment conclusion.

How PAL could be used to inform task design was demonstrated through a system improvement option of adding a testing task. It was shown that adding a testing task that increases the engagement with the primary system could improve attention level and performance in responding to an emergency event. To further examine the generalizability of the model and test hypothesis 3, the model is evaluated in the next chapter in a different task environment where the difficulty in processing the emergency event is varied.

6 Hypothesis 3: Modeling the Impact of Task Difficulty

This chapter describes a human subject experiment data set that was used to test dynamic hypothesis 3. Hypothesis 3 states that with reduced executive control resources and attention on the primary task, human performance on unexpected tasks in low task load supervisory control settings is worse with difficult tasks as compared to easy tasks. The task, experiment design and key results are described. Parameters used in the model are presented. A comparison between model outputs and experimental results is conducted to evaluate the ability of the model to capture the attention and performance in responding to an emergency event with different levels of difficulty. The effect of individual differences is evaluated. A system improvement approach, of adding a secondary task when inattention is detected, is evaluated using the PAL model.

6.1 Experiment Description

With in the increasing level of automation in many supervisory control tasks, human operators typically work under low task load while monitoring these systems, but a rapid transition to high task load situation may occasionally happen when an emergency event cannot be handled by the automation. This experiment investigated the impact of the critical event onset time and the task difficulty on operator performance (Boyer 2014). Thirty individuals participated in the experiment. This section introduces the task, experiment design, and the key results related to testing the model.

6.1.1 Task

The experiment used a simulation that mimics the job of a sensor operator. The operator's job was to monitor the system under normal situation, and track threatening objects if they appeared. The operator needed to reduce track error on simulated threatening objects to a specified threshold. This was a time-pressured task, and as a result, the operator needed to make decisions and perform the actions rapidly. Under normal situations with no threats, the operator did not need to interact with the system other than monitoring and responding to text messages occasionally. Overall, operators worked under low task load under normal situations, but needed to adjust to high task load quickly when threatening objects appeared.

The interface of the simulation is shown in Figure 6-1. Three windows on the left display the three sensor trackers used for tracking objects. In the middle is a window that shows tracking accuracy for one selected object. On the right is a map that shows a 2D representation of the tracking equipment and objects. There is also a system message display, a chat box for communicating with control center, a timer and clock.

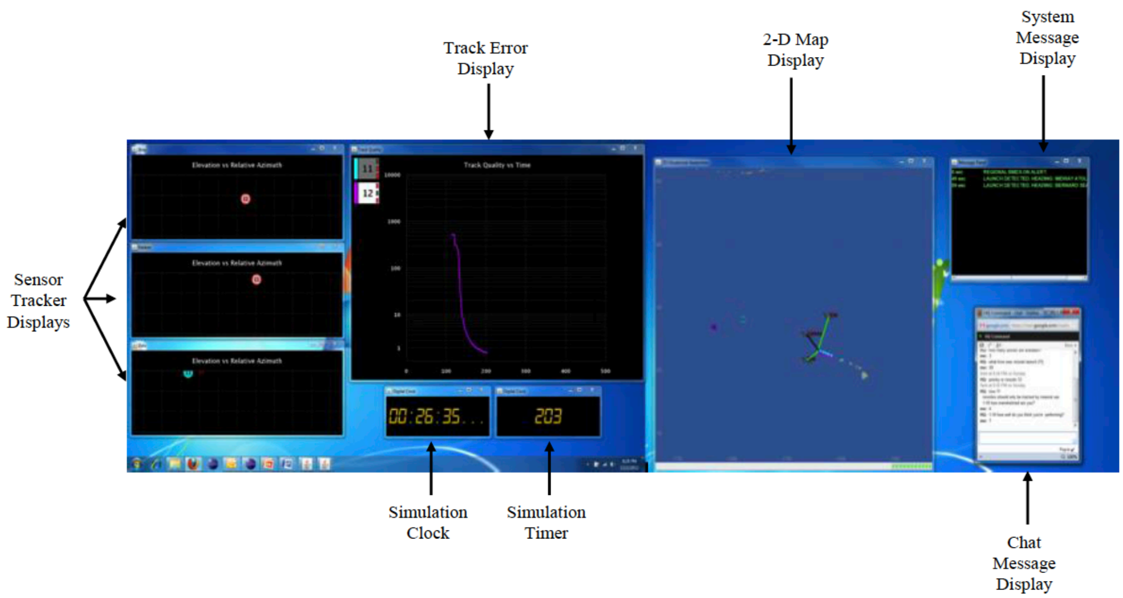


Figure 6-1: Interface of Sensor Tracking Task

The threatening objects follow predetermined trajectories unknown to the operator, and their arrival time is unknown to the operator. The operator receives a message in the System Message Display that the system is on alert before the event, but this message could be a false alarm. Other than that, the operator receives no direct alert before the event. The threatening objects suddenly appear in the Sensor Tracker Windows and in the Tracking Error Display at the start of the event. Once the event starts, the operator has 100 seconds to track the objects. There are three sensors that can be used by the operator. Each sensor is capable of tracking one object at a time. The operator could use sensors together to reduce the track error of a specific object much faster than using only one tracker per object, but then the other objects would not be tracked. During the tracking task, the operator also needs to re-task the trackers to achieve the required accuracy on all the objects.

6.1.2 Experimental Design

The test session of the experiment lasts 180 minutes. There are two independent variables. The first is the event onset time, which has three levels. The threatening objects appear either 40 minutes, 100 minutes or 160 minutes from the start of the mission. Similar to the experiment in Chapter 5, it is expected that the longer people stay in the low task load situation, the worse their performance in responding to the high task load emergency event. The second independent variable is task difficulty. This is varied by changing the number of threatening objects that need to be tracked. In the easy condition, there are three threatening objects. The operator can easily assign one sensor to each object to reduce the track error. In the hard condition, there are six threatening objects. In order to track all six objects, the operator must re-task the three tracking sensors during the mission.

Dependent variables were measured along several dimensions. Performance of the tracking task was measured by percentage of threats tracked to the predetermined track error threshold, which was set based on pilot studies. Absolute final track error was also measured, but was not used in comparison with the model output because the initial track error was unknown in this historical experiment data set. Attention states of the participants were coded based on video recording. Two attention states were used for the coding:

1) Directed: The participants were focused, scanning both displays, or interacting with the interface.
2) Distracted: The participants are drowsy/asleep, looking outside the screen for extended periods, playing with an object besides the display (cell phone, hair tie, etc.), or staring blankly at the screen for a long time without activity.

In the original analysis of the experiment, only attention states 2 minutes before event onset were coded. In order to obtain time series data to compare with model output, all the videos were recoded.

Data on demographics, personality as measured by NEO-FFI 3 score, Boredom Proneness Score, sleep quality, and video game experience were also collected using pre-experiment questionnaires. Boredom Proneness Score was measured by 24 true-false questions (**Error! eference source not found.**). Sleep quality was evaluated by the number of hours slept in the previous two nights. Workload was measured using NASA-TLX and post-event questionnaires after the experiment. Response time to chat box messages were also recorded as secondary workload measurement. Participants' brain oxygenated and deoxygenated hemoglobin concentrations were also measured using a functional near infrared spectroscopy (fNIRS) device. However, these were not used for the comparison with model output.

6.1.3 Results

Performance as measured by percentage of objects tracked was analyzed using nonparametric methods. Task difficulty was shown to be a significant factor (Mann-Whitney $U = 158.5$, $p = 0.042$). Performance under the hard condition was worse as compared to the easy condition, as expected. Onset time was not a significant variable. Examining the data shown in Figure 6-2, when the task was easy, the performance decreased a little with later event onset. When the task was hard, performance with 100 minutes event onset had the worst performance. Performance with 160 minutes event onset was surprisingly better. It was expected that performance would be worse with later event onset time because the long idle time would cause lower executive control resources and distraction. Part of the reason for this might be the small sample size. There were only five participants in each condition. Performance under the hard condition was further analyzed by considering attention states and individual differences.

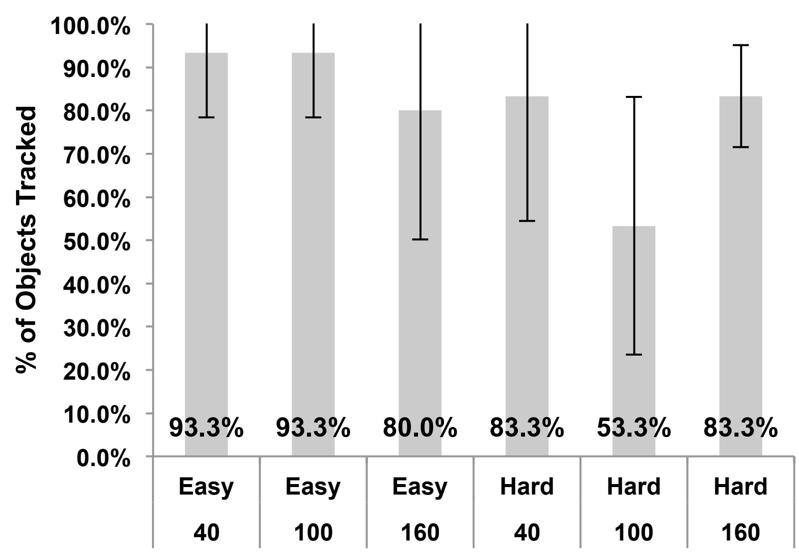


Figure 6-2: Performance as a function of percentage of objects tracked

When directed attention was summarized in 20-minute blocks and used as a dependent variable with repeated measures, it was shown that event onset time has a significant effect on attention ($F(18,198)=2.1496$, $p = 0.0057$). Results from repeated measures ANOVA show that the 100-minute event onset has the lowest attention level, followed by the 40-minute event onset, and the 160-minute event onset. The attention level in the 160-minute event onset condition was surprisingly high. In order to investigate this, the attention levels of the first 20 minutes were examined. The average percentage of time in a directed attention state in the first 20 minutes is listed in the last two columns of Table 6-1. It can be seen that the 160-minute event onset group had a high attention level from the very beginning (Sawin and Scerbo 1995) compared to the other two groups. It was hypothesized that the initial attention level was not affected by the event onset time or the task difficulty. The differences in initial attention level could mostly be attributed to individual differences.

Individual differences on the Boredom Proneness Score and total hours of sleep in previous two nights were analyzed. Among the demographic variables measured, these are the two variables that might impact sustained attention and performance as suggested by previous literature. It has been shown that sleep loss causes difficulty in maintaining sustained attention (Krueger 1989), and high boredom proneness people can lead to worse performance in vigilance tasks comparing to low boredom proneness people (Sawin and Scerbo 1995). In the experimental data, Boredom Proneness Score ranges from 1 to 17, with a mean of 5.76, and standard deviation 3.56. A higher score means that a person is more prone to boredom. In the original study of the Boredom Proneness Scale with 28 true-false items, the average score was 10.44 for males and 9.30 for females (Farmer and Sundberg 1986). Although a 24-item scale was used in our study, the score is lower comparing to the original study.

The average boredom proneness score for each condition is summarized in Table 6-1. In the 100-minute event onset, hard task condition, the average boredom proneness score is 8.2, which is higher than the average score of 4.6 in the 160-minute event onset, hard task condition. Although the hours of sleep in the previous two nights were both low in these two conditions, participants in the 160-minute event onset, hard task condition had a higher average percentage of directed attention (78.72%) compared to 55.63% in the 100-minute event onset, hard task condition. In Figure 6-2, the 160-minute event onset, hard task condition also produced much better performance as compared to the 100-minute event onset, hard task condition. In summary, participants had lower Boredom Proneness Scores, and higher levels of directed attention in the 160-minute event onset, hard task condition, which may have contributed to their better performance. Regression analysis was run to test the impact of sleep and boredom proneness on attention. Unfortunately, no significance was found. The impact of these two variables on attention needs to be investigated with additional data. In addition, other factors such as fatigue that were not measured in this study might influence the attention level as well.

Table 6-1: Descriptive Statistics of Dependent Variables

Event Onset	Task Difficulty	Sleep		Boredom Proneness		Directed Attention: Overall (%)		Directed Attention: First 20 min (%)	
		mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
40	Easy	16.2	2.77	5.2	3.83	74.10	19.33	90.85	3.84
40	Hard	16.1	2.30	4.4	1.34	54.62	20.23	69.12	22.44
100	Easy	14.0	0.707	6.6	3.51	54.73	36.51	70.60	35.98
100	Hard	13.5	0.707	8.2	6.10	55.63	22.68	75.17	17.03
160	Easy	16.0	2.345	5.6	3.13	58.68	24.24	77.58	14.58
160	Hard	12.6	2.302	4.6	1.82	78.72	13.32	90.60	3.39

In summary, the experiment found that the difficulty of the emergency event has a significant impact on performance. The impact of emergency event onset time on performance was not significant, which may have result from the small sample size and individual differences on boredom proneness and sleep quality. Comparing this experiment with the one in Chapter 5, although both include a shift from low task load to high task load, the task characteristics are quite different. The emergency event in this experiment needed to be processed within 100 seconds, while for nuclear power operator experiment, the alarm took about 51.21 minutes to be cleared. In addition, tracking objects needs vigilance, fast response, and eye-hand coordination. Clearing the alarm in a nuclear power plant, however, needs more understanding of the system and problem-solving skills. These differences influence the dynamics of attention during the mission and how attention would influence performance.

6.2 Model Parameters

In order to test whether the system dynamics model in Chapter 3 could capture the change of human attention and performance in responding to emergency events with different difficulty levels, the model outputs were compared with the experiment data. The parameters used in the model are shown in Table 6-2. *Baseline Self-Imposed Event Rate* was set to zero in this case, as extra interactions with the system are negligible. Under normal conditions, the operators only need to respond to chat messages occasionally.

Table 6-2: Model Parameters

Model Parameters	Values	Other Parameters
Baseline Self-Imposed Event Rate	0 tasks/minute	$k_1 = 9.39$
Exogenous Event Rate	PULSE(Pulse Time, Pulse Size)	$k_2 = 1.99$
	Pulse Time = 40, 100 or 160	$c_1 = 2.05$
	Pulse Size = 3 or 6	$c_2 = 10$
Normal Processing Rate	3 tasks/minute	$k_3 = 8.96$
Required Processing Time	2 min	$k_4 = 2.54$
Time Window	5 min	$k_5 = 5$
Depletion Time	180 min	$c_5 = 9.11$
Restore Time	180 min	$m_5 = 0.04$
Attention on Task, Initial value	1	$k_6 = 10$
Average Time to Distract	30 min	$c_6 = 5.49$
Refocus Time	1 min	$k_7 = 10$
Executive Control Resource:		$c_7 = 1.20$
Initial value	Personal Precursors	$b_0 = -0.60$
Maximum Level	Personal Precursors	$b_1 = 0.79$
Minimum Level	0	$b_2 = 0.82$

Exogenous Event Rate was set based on the event onset time in the experiment. The PULSE function in Vensim® was used to model the arrival of one emergency event at the onset time at 40 minutes, 100 minutes or 160 minutes since the start of the experiment. The number of objects to be tracked was modeled using variable *Pulse Size*, which equaled to either three or six as in the experiment setting. *Required Processing Time* was set for 2 minutes, which is rounded from the required 100 seconds in the experiment. The task was designed so that it is possible but challenging to track all the objects within the required 100 seconds. Dividing 6 objects by 100 seconds equals to 3.6 tasks/minute. Thus, *Normal Processing Rate* equals 3 tasks/minute. This results in a total of 6 objects tracked within the 2-minute

Required Processing Time. *Average Processing Rate* was calculated as an exponential moving average of *Event Processing Rate*. To reflect the recent changes in task load, the sampling time interval for the moving average (*Time Window*) was set at 5 minutes.

For *Executive Control Resource*, both the *Depletion Time* and *Restore Time* are set to 180 minutes, which is the length of the mission in the experiment. For attention, *Attention on Task* is initialized at 1, because participants paid full attention on task at the beginning of the experiment. *Average Time to Distract* is 30 minutes, because previous research shows that sustained attention decreases around 20-30 minutes into the mission (Mackworth 1957). *Refocus Time* was set to 1 minute, because participants needed to adjust their attention within the 100 seconds task processing time in this experiment. This means participants increase their attention after event onset, and reach the maximum level of attention in 1 minute.

The minimum level of *Executive Control Resource* equals zero. The initial value and the maximum level of *Executive Control Resource* are set based on *Personal Precursors*, which follows Equation (38). Research shows that sleep deprivation impairs executive control (Martella et al. 2011). For each condition in the experiment, the average amount of sleep was calculated and used in the model. In the experiment data, the maximum amount of sleep for the previous two nights was 21 hours. Thus, the average total hours of sleep in the previous two nights for each condition was divided by 21 to get a value between zero and one. In this equation, more sleep means a higher value for *Personal Precursors*, which leads to a higher level of maximum executive control resource, and vice versa. Boredom proneness follows the opposite relation. People who are more prone to boredom have lower levels of maximum executive control resource.

For each condition in the experiment, the average boredom proneness score was calculated and used in the model, as listed in Table 6-1. Since boredom proneness was measured by a 24-item true-false questionnaire in this experiment, the score was divided by 24 to normalize it. *Personal Precursors* also impacts *Time to Distract*. People who have less sleep or more prone to boredom get distracted faster, which means a shorter *Time to Distract*. The impact of individual differences on *Time to Distract* was modeled in Equation (39). *Average Time to*

Distract was a constant value. The actual *Time to Distract* was varied based on this constant value by multiplying it with *Personal Precursors*.

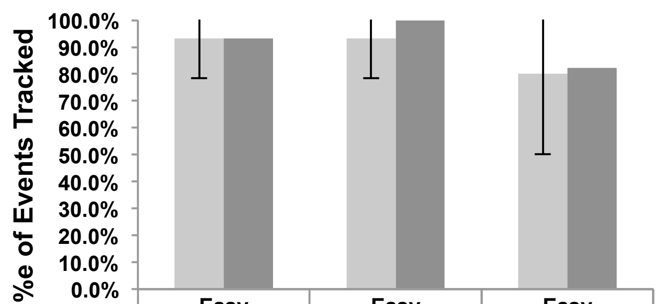
$$Personal\ Precursors = b_0 + b_1 * MAX\left(\frac{Sleep}{21}, 1\right) + b_2 * \left(1 - \frac{Boredom\ Proneness}{24}\right) \tag{38}$$

$$Time\ to\ Distract = \ Average\ Time\ to\ Distract * Personal\ Precursors \tag{39}$$

The coefficients b_0, b_1 and b_2, as well as other variables listed in the last column in Table 6-2 were calibrated using Vensim®. The nonlinear relationships were visualized in Appendix A. The objective of the model calibration was to find a set of parameter values that provides the best match between model outputs and experiment data on performance and attention. Instead of trying with thousands of combinations of different parameter values, the calibration process was automatically executed by Vensim®. Experiment data with the easy task was used for model calibration, and the data with the hard task was used for prediction validation.

6.3 Model Fit

In order to test whether the model could successfully capture human performance and attention in responding to an emergency event that is easy to handle, two key outputs of the model are compared with the experiment data. The model output on performance in terms of percentage of objects tracked with sufficient accuracy is compared with experiment data. The mean and standard deviation of performance in the experiment are shown in Figure 6-3. Overall, the model provides a good estimation on performance when the task was easy.



	Easy 40	Easy 100	Easy 160
Experiment	93.3%	93.3%	80.0%
Model	93.2%	100.0%	82.3%

Figure 6-3: Performance of Easy Task

Table 6-3: Simulation to Experimental Data Fit Statistics (Easy Task)

Summary Statistics	Attention on Task (Easy Task)		
	Early Onset	Middle Onset	Late Onset
Coefficient of Determination (R^2)	0.248	0.142	0.280
Root Mean Square Error (RMSE)	0.135	0.127	0.131
Mean Square Error (MSE)	0.018	0.016	0.017
Bias component of MSE (U^M)	0.163	0.002	0.252
Variation component of MSE (U^S)	0.108	0.071	0.180
Covariation component of MSE (U^C)	0.728	0.927	0.568

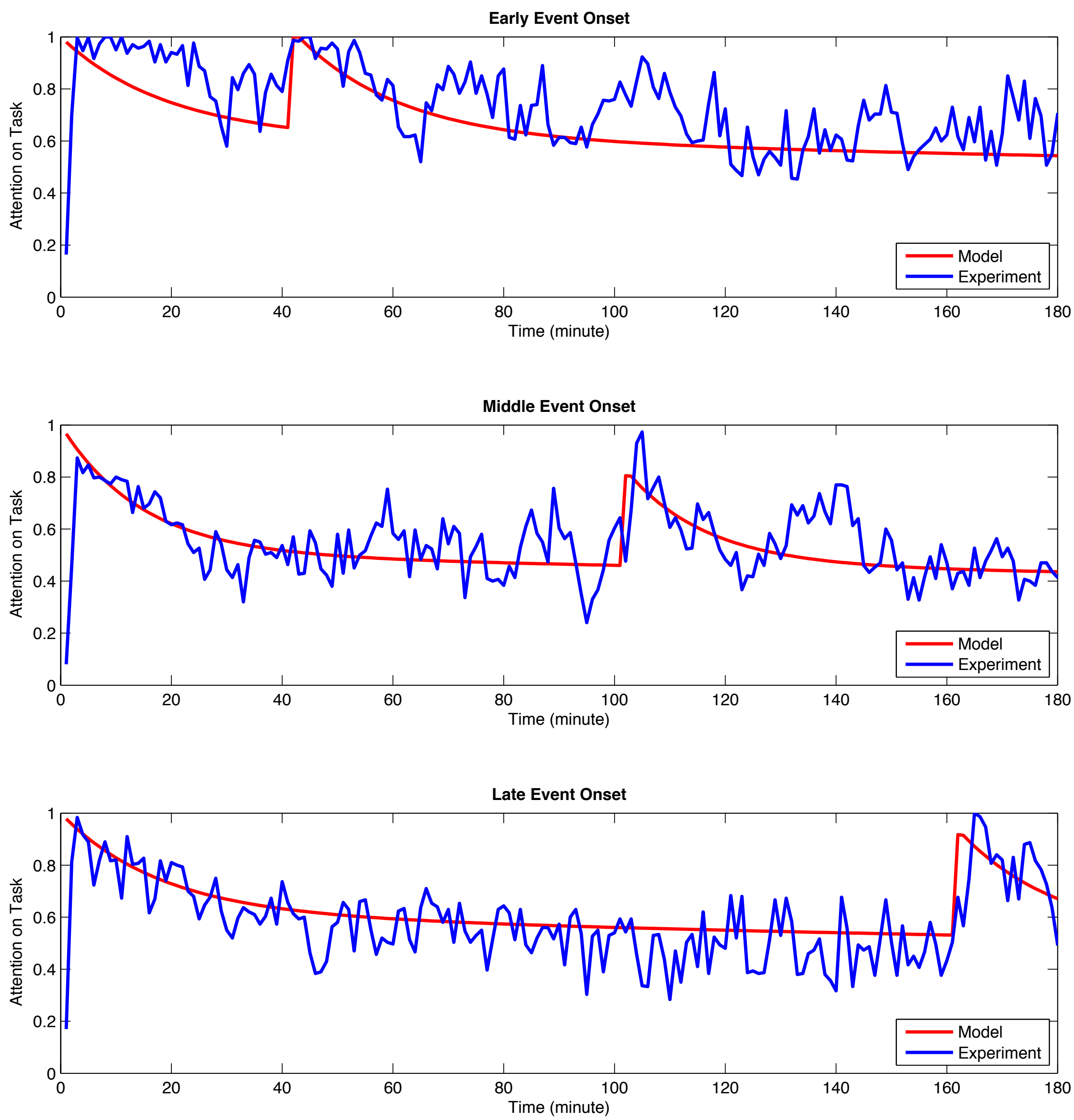


Figure 6-4: Attention on Task with the Easy Task

Attention on Task is compared with the average level of directed attention in the experiment data, which is presented in blue lines in Figure 6-4. A quantitative assessment of the model's fit to experimental data with easy task is provided in Table 6-3. Percentage of time in the directed attention state is summarized by minute, as shown in the blue line in Figure 6-4. The red line shows the change of *Attention on Task* in the system dynamics model. In general, the model captured the decrease of attention under normal conditions and the increase of attention with emergency event onset. Since the task requires only about two minutes to process, attention on task increases as the emergency event happens, and decreases quickly as the task is completed. The level of attention is also impacted by sleep and boredom proneness as captured by *Personal Precursors*. Average hours of sleep and boredom proneness scores of each condition as shown in Table 6-1 were used in the model. The middle onset, easy task condition has a lower level of attention partly because people in this group had less sleep. This is correctly reflected in the model output.

Overall, the R^2 values are low. It is affected by the fact that small variations of attention are not captured in the model. For all three groups, U^C is the largest component of MSE, indicating a good fit. The small values of U^M and U^S indicate small errors due to bias and unequal variation. Since the purpose of the model is to capture the general trend of attention change rather than to predict point-to-point attention state, the current model fit is good enough for this purpose.

6.4 Predicting the Impact of Task Difficulty

In order to evaluate the model's ability to predict the impact of task difficulty, the model outputs for attention and performance with the hard task were compared with the experimental data. Since only data with the easy task was used for model calibration, this provides a good prediction validation. All the parameters calibrated using data with easy task were kept the same. The value of *Pulse Size* was changed from three to six to reflect the impact of task difficulty.

The quantitative statistics of model fit are presented in Table 5-1. Overall, R^2 values are not very good. U^C was the largest component of MSE under the middle onset condition, which indicates a good model fit. Under both early and late onset conditions, there is a bias in prediction as reflected by the value of U^M. Visually observing the change of attention as shown in Figure 6-5, the model over-predicted the level of attention in early onset condition, and under-predicted the level of attention in late onset condition.

Table 6-4: Simulation to Experimental Data Fit Statistics (Hard Task)

Summary Statistics	Attention on Task (Hard Task)		
	Early Onset	Middle Onset	Late Onset
Coefficient of Determination (R^2)	-0.540	0.114	-2.491
Root Mean Square Error (RMSE)	0.213	0.161	0.262
Mean Square Error (MSE)	0.046	0.026	0.069
Bias component of MSE (U^M)	0.603	0.192	0.747
Variation component of MSE (U^S)	0.093	0.168	0.020
Covariation component of MSE (U^C)	0.304	0.639	0.232

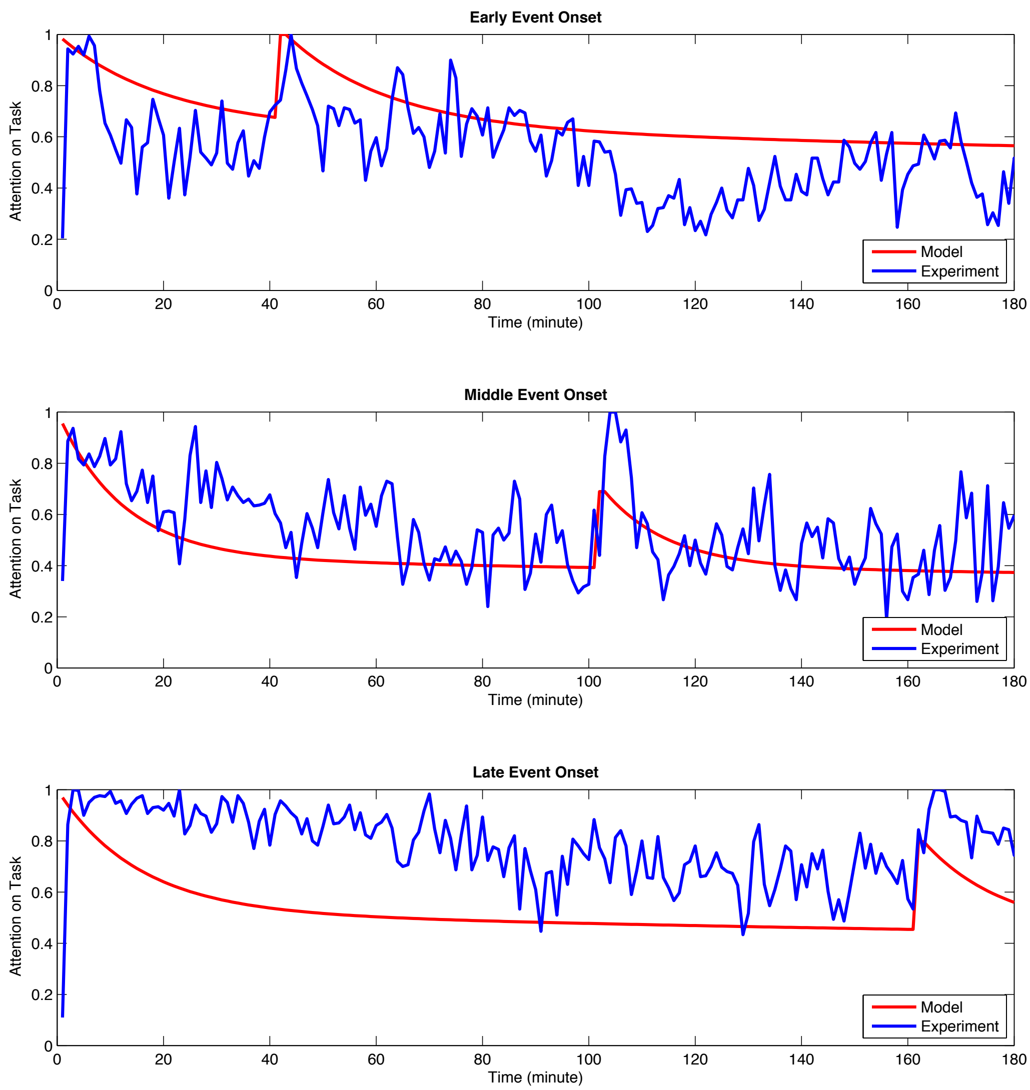
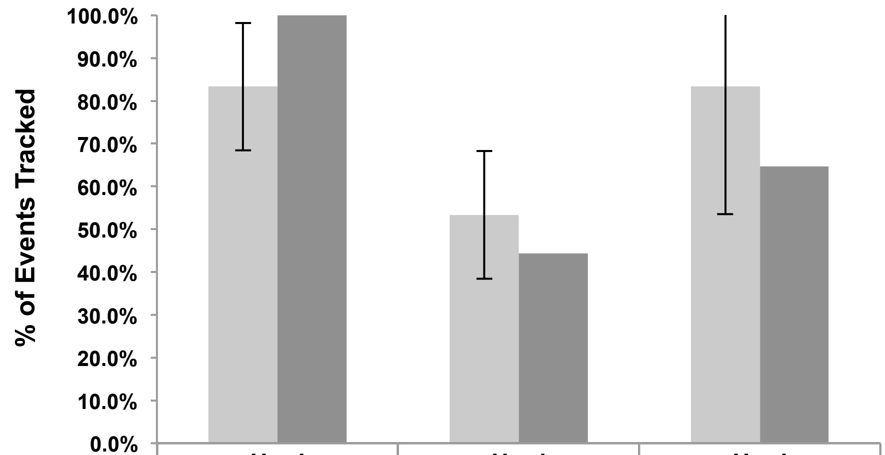


Figure 6-5: Attention on Task with the Hard Task

Although it is normal to have a worse fit in prediction compared to model calibration, this reflects that there might be something missing in the model. In the experimental data, the early onset easy task group had 16.2 hours of sleep on average comparing to 16.1 hours of sleep in early onset hard task group. The early onset easy task group had a boredom proneness score of 5.2 on average, comparing to 4.4 in the early onset hard task group. The model predicts similar attention levels based on these data. However, the actual attention level in the early onset hard task group is lower than the early onset easy task group. There are a few possible reasons. First, there are additional variables related to the individual differences on executive control resource, which are not captured in the model, such as motivation (Inzlicht et al. 2014). In the experiment, participants were told that the best performer would be awarded a $150 gift card in addition to the $75 compensation for participating the experiment. It is possible that participants were more motivated than others in getting this award, thus maintaining a high level of executive control and sustained attention. Second, the impact of sleep and boredom proneness on executive control resources may not be linear. It is possible that executive control resource would decrease a little bit when there is a little sleep loss, but more rapidly with a higher level of sleep loss. Third, the amount of sleep each person needs is likely to be different (Van Dongen and Belenky 2009). Some people are more vulnerable to sleep loss. Lastly, the small sample size for each condition may create an experiment artifact. There are only five participants in each condition.

The bias in prediction on attention also impacts the prediction on performance. As shown in Figure 6-6, the model under-predicts performance in the early onset hard task condition. The model also over-predicts performance in the early onset hard task condition. Despite the bias on attention and performance, the model did predict a decrease of performance when the task difficulty was increased. This is consistent with the experiment conclusion. Performance with easy task did not decrease much when the emergency event happened later. However, when the task was hard, the decrease of performance is quite obvious in both middle and late onset groups. In dynamic hypothesis 3, it states that human performance on unexpected tasks is worse with difficult tasks as compared to easy tasks when executive control resource and attention on the primary task were decreased. This is supported by the experiment and model results.



	Hard 40	Hard 100	Hard 160
Experiment	83.3%	53.3%	83.3%
Model	100.0%	44.3%	64.7%

Figure 6-6: Performance with the Hard Task

6.5 Impact of Individual Differences

In the previous model testing process, mean values of sleep and boredom proneness were used for each condition. Sleep and boredom proneness scores were used to calculate the value of *Personal Precursors* in the model. In order to understand the impact of individual differences, extreme values of *Personal Precursors* based on experiment data were used. Based on Equation (40), values of the variable *Personal Precursors* were calculated for all participants. The coefficients used in this equation were calculated as shown in Table 6-2.

$$\begin{aligned} &Personal\ Precursors \\ &= -0.60 + 0.79 * MAX\left(\frac{Sleep}{21}, 1\right) + 0.82 * \left(1 - \frac{Boredom\ Proneness}{24}\right) \end{aligned} \tag{40}$$

Given the experimental data, the maximum and the minimum values of Personal Precursors are 0.128 and 0.832. The corresponding values for sleep and boredom proneness score are listed in Table 6-5.

Table 6-5: Extreme Values for Personal Precursors

	Sleep	Boredom Proneness Score (BPS)	Personal Precursors
Low Sleep, High BPS	13	17	0.128
High Sleep, Low BPS	19	3	0.832

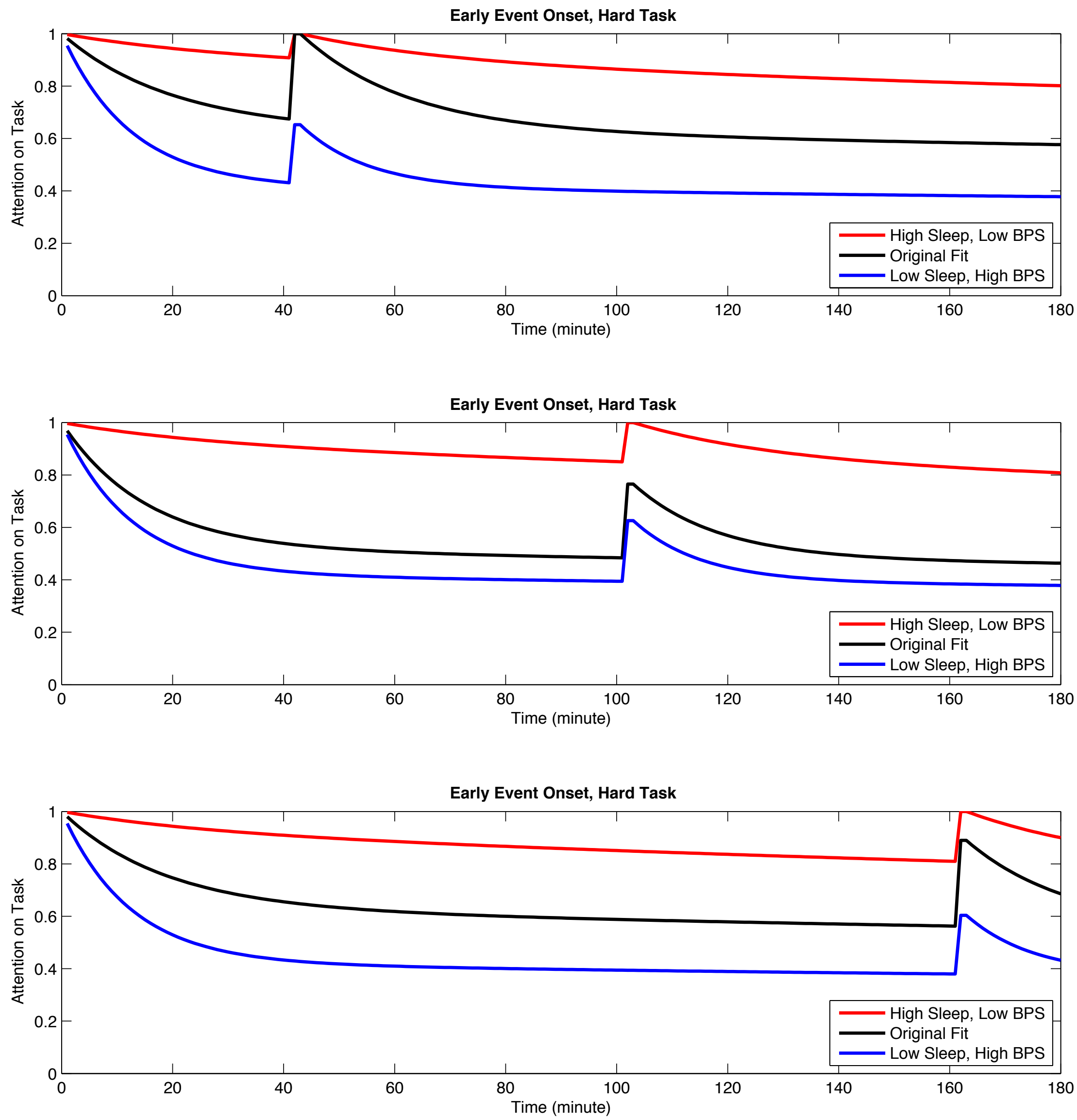


Figure 6-7: Impact on Attention with Easy Task for Different Personal Precursors Values

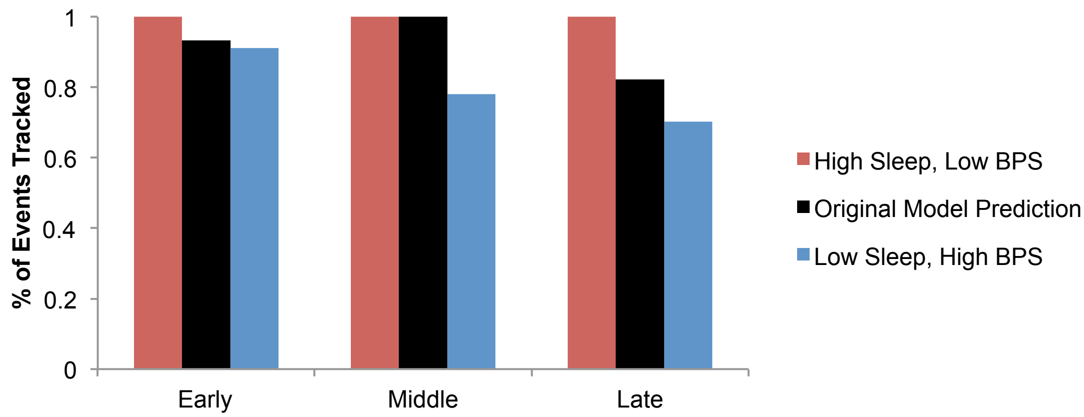


Figure 6-8: Impact of Personal Precursors on Performance with Easy Task

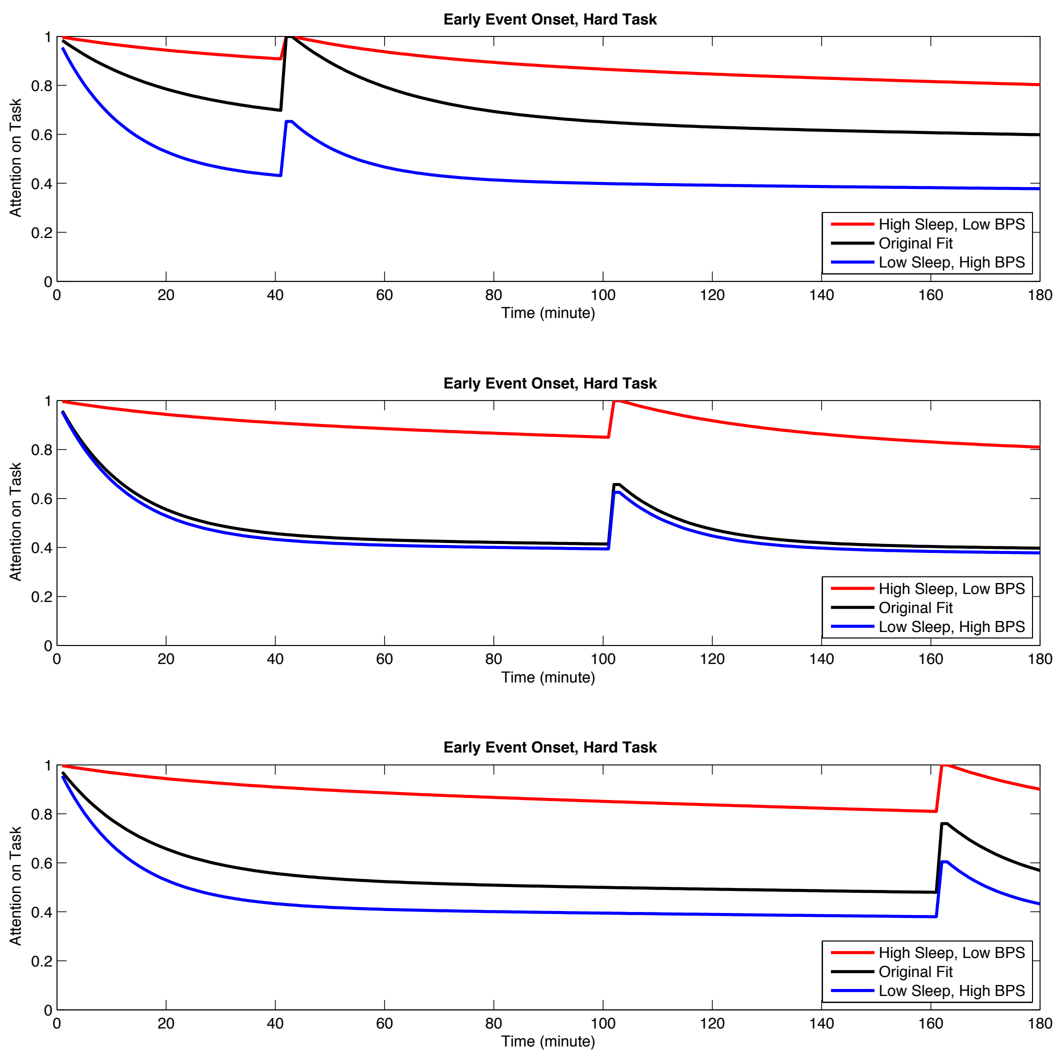


Figure 6-9: Impact of Personal Precursors on Attention with Hard Task

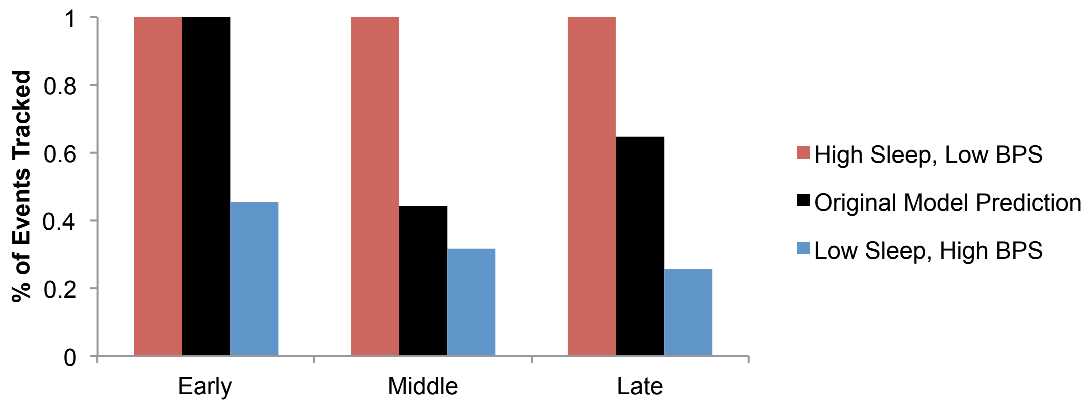


Figure 6-10: Impact of Personal Precursors on Performance with Hard Task

The impact of such changes on attention and performance with easy task are shown in Figure 6-7 and Figure 6-8. The original outputs of the model are shown in black lines. The red lines show the outputs with more sleep and low boredom proneness, and blue lines show the outputs with less sleep and high boredom proneness. The graph shows that attention can be improved or decreased depending on the value of sleep and boredom proneness. For performance, Figure 6-8 shows the number of objects tracked with the easy task. It can be seen that the improvement on performance is limited if the performance is already relatively high. If the original performance is worse, the improvement can be larger.

The impact of personal precursors on attention and performance with hard task is shown in Figure 6-9 and Figure 6-10. The original outputs of the model are shown in black lines. The red lines show the outputs with more sleep and low boredom proneness, and blue lines show the outputs with low less sleep and high boredom proneness. Similar to the easy task, the graph shows that attention can be improved or decreased depending on the value of sleep and boredom proneness. For performance, Figure 6-10 shows the number of objects tracked with the hard task. It can be seen that the improvement in performance is much higher with more sleep and low boredom proneness score comparing to the cases with the easy task.

In conclusion, the overall level of attention is higher when the people have more sleep and are less prone to boredom. The largest improvement on performance was under the 100-

minute event onset, hard task, in which all six objects were tracked (100% of all objects) with sufficient accuracy comparing to 2.658 (44.3% of all objects) in the original case. When the value of *Personal Precursors* was low, which means people have less sleep and are more prone to boredom, attention decreases much faster. In the condition with the 100-minute event onset and hard task, 31.7% of the objects were tracked with sufficient accuracy with less sleep and more boredom proneness. This analysis shows that performance in responding to unexpected emergency event can be improved by using personnel with low boredom proneness scores and sufficient sleep. A further analysis with different combinations of sleep and boredom proneness score was conducted as described in Appendix F to assess their impact on task performance. With the comparison with experimental data, it shows boredom proneness is less influential than sleep.

6.6 System Change Prediction: Task Processing Capability

In previous sections, it was shown that the decrease of attention over time during the low task load monitoring period leads to degraded performance in responding to an emergency event, especially for the hard task. One approach to addressing this is to improve attention management, as shown in Section 5.5, where a secondary testing task was added. While this could improve performance in responding to an emergency event, it is not sufficient to address the challenge brought by difficult tasks. If the emergency event is too difficult to handle, the operator may not achieve good performance even with full attention. In this section, the impact of task processing capability is investigated using the PAL model. Task processing capability here refers to how fast the task can be processed with full attention and full ECR, which is represented by *Normal Processing Rate* in the model. For the task scenario discussed in this chapter, *Normal Processing Rate* in tracking threatening objects is affected by several factors, including the number of sensors, system interface design, automation level in sensor management, as well as individual capability and strategy.

The hard task was used for testing, which had six objects to track. The value of *Normal Processing Rate* was varied from 1 to 6 tasks/minute to assess its impact on performance. The baseline value was 3 tasks/minute, corresponding to the experimental data. In the experiment, the hard task required the operator to track six objects in about two minutes. All the other parameters were kept unchanged.

The impact of changing *Normal Processing Rate* on performance is presented in Figure 6-11. Performance is improved with a higher *Normal Processing Rate* but differs among the three event onset times. For example, a *Normal Processing Rate* of 3 was sufficient to track all six objects in the early onset condition. However, in the middle onset condition, even 6 tasks/minute was not enough to track all the objects. If a certain performance goal was set, system designers could use the PAL model to assess how fast the task processing capability should be to achieve the desired goal.

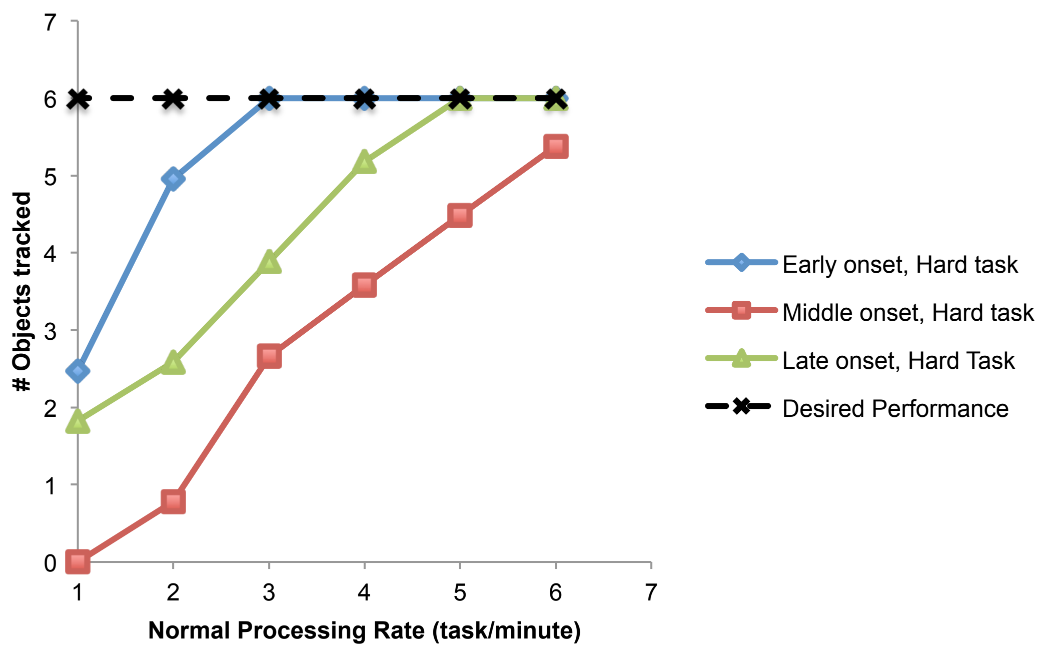


Figure 6-11: Impact of Normal Processing Rate on Performance

The improvement in performance differed across the three event onset times. In fact, it reflected a more fundamental difference on the attention and ECR level prior to event onset, which were affected by onset time and individual differences. For example, the middle onset condition had the worst performance overall despite the increase in *Normal Processing Rate*. This is because it had the lowest attention and ECR level, as presented in Figure 6-12. Attention was not affected by *Normal Processing Rate* in the PAL model. In other words, no matter how fast a person could process an emergency event, attention on primary task decreased at the same rate as during normal conditions when nothing happens. It can be argued that people may be more motivated to maintain their attention if they perceive the task to be challenging. However, emergency events happen unexpectedly, which means people have no prior knowledge of when it will happen and how difficult the task will be. This independence between attention and task processing capability means that system designers could use the PAL model to prioritize system improvement options. For example, if task processing capability cannot be increased due to limited resources (i.e., cannot increase along the performance curve), attention management should be the focus for improvement (shifting the curve up in Figure 6.12).

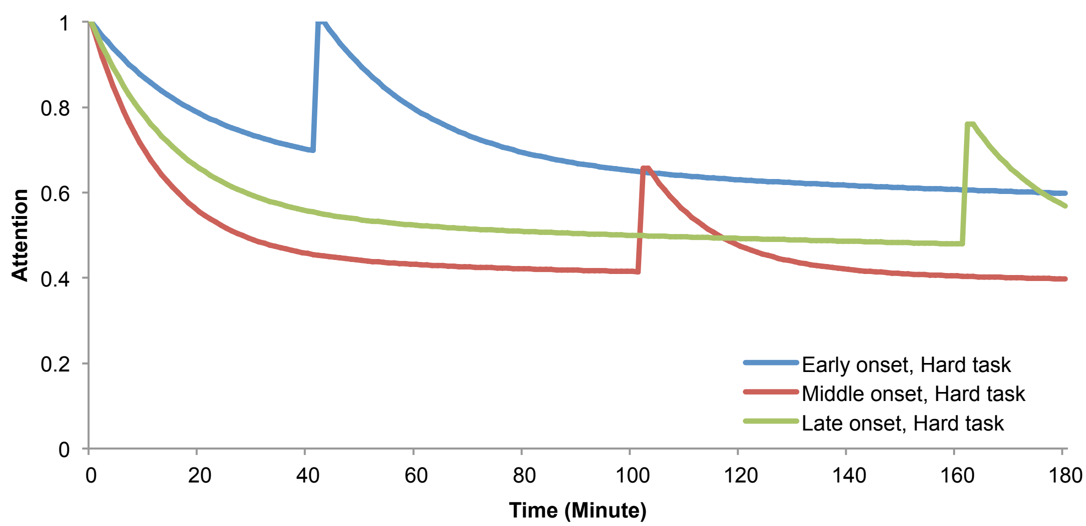


Figure 6-12: Change of Attention with Different Event Onset Times

In summary, performance in responding to emergency events could be improved if task processing capability is increased. These increases in performance follow different curves as affected by attention and ECR level. The PAL model could facilitate system designers to assess the impact of increasing task processing capability on performance and prioritize between improving attention management and increasing task processing capability.

6.7 Chapter Summary

In this chapter, a long duration human subject experiment with a shift from low task load to high task load was introduced. In this experiment, participants monitored the control interface of sensor trackers and responded to an emergency event by tracking the threatening objects. Three levels of event onset time were tested to investigate whether the duration of monitoring time influenced operator performance while responding to the emergency event. The difficulty of task was also varied by the number of objects needed to be tracked. The system required only infrequent human interactions under normal conditions, but high effort in a very short time when the emergency event happened. In the experiment, the performance with a hard task was worse as compared to an easy task. Moreover, it seems the performance with the hard task was more influenced by event onset times, although statistical testing showed no significance.

In order to assess the ability of the model to replicate these behaviors, model parameters were set based on experiment data, previous literature, and model calibration using part of the experiment data (data with easy task). The outputs of the model on attention and performance were compared with the experiment data. The comparison shows that the model could successfully replicate the experimental behavior in attention change and performance with the easy task.

To evaluate the model's ability to predict the impact of task difficulty, model predictions on performance and attention were compared against experiment results with the hard task. Despite some bias, the model predicted an overall decrease in performance when the task difficulty was increased, supporting dynamic hypothesis 3. The influences of individual differences as reflected in sleep and boredom proneness were also captured in the model. The impact of individual differences was evaluated using two extreme values. The impact of task processing capability on performance was evaluated using the model. This demonstrated how the PAL model could facilitate the system design process.

These results further confirm the validity and usefulness of the model. Although the response to an emergency event was evaluated in both Chapters 5 and 6, these two tasks have many differences in terms of the time required to process, the required ability, and the type of interactions. These differences were represented in the changes of parameter values with the model structure unchanged. In addition, there were six experimental conditions in both Chapters 5 and 6, under which the behaviors were correctly captured. In Chapter 4, the model represents a low task load case with no emergency event. The model needed only a little modification to represent the actual performance without changing the main feedback structures. This shows that the PAL model is able to represent the behavior of different human supervisory control tasks using the same model structure. In other words, it is a general theory of how operators behave under low task load, rather than a model tailored to a specific task scenario. The fact that the model could generate the behavior in different task settings shows the generalizability of the model.

7 Conclusions

With the move towards more automated systems in the future, the issues with low task loading in automated environments require more attention. A better understanding is needed to enable intervention and mitigation of possible negative impacts. In order to systematically understand human behavior and performance in low task load automated environments, a system dynamics model, the Performance and Attention under Low-task-loading (PAL) Model, was developed to capture the dynamic changes of attention and performance impact when interacting with automated systems with low task load. This model can be used to facilitate system design and evaluate different design options. Three dynamic hypotheses were developed and tested with three experimental data sets under different scenarios.

7.1 The PAL Model

In supervisory control tasks, the increase in automation level usually means a decrease of human interaction. Often a development in automation technology creates a low task load scenario, in which humans do not need to do much work other than monitoring the system, and only intervening when a problem occurs. However, this system design approach introduces new concerns, as it is difficult for humans to maintain sustained attention, especially in an environment with little stimuli. In fact, the decrease of attention on the task and redirection of attention to distractions is almost inevitable under such low task loading.

While the negative impacts of repetitive tasks have been investigated, the understanding of the effects of low task loading in the presence of significant automation is limited. People often regard the decrease of task load as a benefit and try to increase the automation as much as possible to achieve this. While automation can bring huge benefits in improving efficiency and reducing workload for complex and difficulty tasks, low task load can also be dangerous. The low task load could cause mind wandering and distraction in attention and the loss of situation awareness as well as boredom and frustration. These could decrease the overall system performance especially when there is an event beyond the automation capability. In the long run, the reliance on automation can also cause skill degradation. The erosion in manual flying skills is a contributor to many accidents (Geiselman et al. 2013). In order to investigate human performance and attention change under low task load scenario, a conceptual framework called the Boredom Influence Diagram (BID) was proposed. It includes components including task characteristics, boredom, fatigue, attention lapse, attention management strategies and performance.

Building on the BID, a system dynamics model, Performance and Attention with Low-task-loading (PAL) model was built to capture the dynamic changes of these components. It has five modules: Task Characteristics, Processing and Performance; Stress; Workload; Executive Control; and Attention Management. Six major feedback loops were formed by these modules: Yerkes-Dodson Loops, Ego-Depletion Recovery Loop, Refocus Loop, Drained from Boredom Loop, Attention Control Loop, and Increase Task Engagement Loop. The six feedback loops capture different aspect of human-automation interaction under low task load. Based on the model results, the two most important loops are Drained from Boredom Loop and Attention Control Loop. As shown in Chapter 4-6, these two loops together cause the decrease of attention over time under low task load, which further influences task performance. These feedback loops not only explain the behavior of human operators under low task load, they also provide guidance for exploring system improvement options. The inputs, outputs, and validity of the model are discussed in the following sections.

7.1.1 Model Inputs and Outputs

The purpose of PAL is to provide a mechanism for system designers to investigate human behavior and facilitate system design. More specifically, the exogenous input parameters allow users of PAL the ability to explore the design space in three categories: system characteristics, human-automation interaction attributes, and human characteristics.

Table 7-1: System Characteristics

Task Scenario	Task Frequency	Task Difficulty	Task Processing Time
Human-Automation Collaborative Searching	1/20 minutes		3 seconds
Nuclear Power Plant Monitoring	1/4 hours Different onset times		About 60 minutes
Threaten Objects Tracking	1/3 hours Different onset times	3 or 6 objects	100 seconds

7.1.1.1 System Characteristics

The first aspect is system characteristics. A core feature of PAL is that it allows designers to manipulate system characteristics in order to represent different supervisory control systems. While it is not possible to represent the details of these tasks in the model, the main characteristics regarding task arrival and task processing can be captured in PAL using a few variables. In Chapters 4-6, three distinct task scenarios were modeled. Their characteristics are summarized in Table 7-1.

These system characteristics are influenced by specific task environment and system design.

- *Task Environment*. For example, clearing an alarm in anuclear power plant requires vigilance as well as complex problem solving. As a result, it took longer to complete a task. For tracking threatening objects, the task had to be completed within 100 seconds to avoid disastrous consequences. The length of time that a task must be completed in and the length of time an individual needs to process a task are modeled as input parameters in the model. Other factors related to the task environment include number of targets to search, frequency and timing of emergency events, etc. While these are not controlled by the system designer for descriptive modeling, the potential impact of these factors could be evaluated using the model for risk assessment or exploring robust design.
- *Degrees of Automation*. While there is research proposed that automation should be categorized by information processing stages in addition to degrees of automation (Onnasch et al. 2014), a simple assumption is that higher levels of automation reduce the frequency of task requests and the time it takes to process a task, which is captured by task arrival process and task processing time in the model.
- *System Design*. In previous chapters, a few system design options were evaluated using the model, including attention alerts, restricting external distraction sources, adding a testing task, and changing task processing capability. Different from previous two categories, system designers have more freedom in representing these changes in the model. Sometimes small changes to the model structure and the formulas are necessary to account for these design changes. As already demonstrated, PAL could be used to evaluate the impact of these design changes on attention and performance.

7.1.1.2 Human-Automation Interaction Attributes
Human-automation interaction attributes refer to the style or strategy of an individual operator in interacting with an automated system. For example, how often an operator creates extra tasks in addition to system requirements during low task load periods can be captured in PAL. Although whether the operator could do so may be restricted by system design, an added active interaction may result in better attention management and performance compared to passive monitoring. This is demonstrated in Chapter 4 in the part on increasing task engagement.

7.1.1.3 Human Characteristics
The third category is human characteristics. The sleep quality and boredom proneness score of individual operators are modeled as input parameters. They are assumed to influence executive control resources and time to distraction. In Chapter 6, the impacts of changing these variables were explored using two extreme cases. While not included in the model, gaming experience, personality and other individual characteristics can also be tested by incorporating them in the variable *Personal Precursors*. However, to do this, causal relationships from literature or pilot studies are required.

The main outputs of the PAL are attention on task and performance. Measuring attention is important especially in low task load environments because doing so could provide information on operator status even when there is no observable event, in which case there is no performance measure. The inputs listed above can be evaluated based on the outputs generated from the model. As shown in Chapters 4-6, the impact of low task loading, event onset times, and task difficulty were evaluated using model outputs on attention and performance. As demonstrated by the predictive case studies in Chapters 4-6, PAL can be used to explore design spaces, both in terms of human performance and system attributes, in order to improve operator attention and both operator and system performance.

7.1.2 Model Validity
PAL was designed based on previous studies on boredom, fatigue, workload, attention and performance. The major causal links in the model are all supported by theoretical or empirical results in literature. The model structure was modified through several iterations by reviewing with experts, as described in Appendix B. The model was further evaluated by examining the model boundary and dimensional consistency. Subsystems of the model were

isolated and tested by adding one loop at a time. This ensures that each subsystem generates reasonable behavior. In addition, discrete and continuous task arrival processes were compared.

Three dynamic hypotheses were proposed, each building on the previous one and expanding the scope of behaviors that can be explained using the model step by step. The first hypothesis stated that attention decreased under low task load. By comparing model outputs with experiment data, it was shown that the PAL model could capture this behavior. In the model, such decreases in attention were assumed to be the result of depleted executive control resources. This causal link was supported by previous literature, but lacking empirical data. It is very difficult to measure executive control resource level, which cannot be measured directly by current technology. Future research regarding this causal link is needed to enhance the validity of the model.

Building on Hypothesis 1, the second hypothesis and third hypothesis examined the impact of decreased attention on performance in responding to an emergency event. This represents a shift from low task load to high task load during a single mission. Both the experiment and the model outputs showed that performance is worse with lower attention levels. The impact of decreased attention on performance was larger when the task was difficult. The process of testing these three hypotheses shows that the PAL model could explain behaviors under low task loading as observed in experimental data, which helps to build confidence in the validity of this theory. The PAL system dynamics model can be regarded as representing a theory to explain certain behaviors in low task loading environments.

Using three experimental data sets, the PAL model was tested to build confidence in the accuracy of the model in different task scenarios. The first experiment involved a search task under low task load with no emergency event. The second experiment represented a nuclear power plant monitoring task in which an emergency alarm needed to be handled. In the third experiment, operators need to track different numbers of threatening targets that suddenly appeared. In all three cases, the PAL model could successfully capture the change in attention and performance. What's more, only part of the experimental data from each experiment was used for model calibration. The predictive power of the model was tested using the rest of the experimental data. The prediction validation shows that the model does not overfit the experiment data. More important, it demonstrated that system design changes and interventions could be evaluated using the PAL model. The diversity in task characteristics shows that PAL can be applied in different settings in low task loading environments. Three distinct task scenarios were represented by changing parameter values with little modification to the model structure. In other words, the PAL model is a general theory of how operators behave under low task load, rather than a model tailored to a specific task scenario. While such a model can never be truly validated, the theoretical foundation, structural tests, and the comparison between experiment data and model outputs help to build confidence in the model.

7.2 Model Generalizability and Limitations

One of the aims of this research is to build a generalized model for human automation interaction in low task load environment. In the three experimental data sets used for model testing, the tasks are very different. The success in replicating and predicting human attention and performance in three distinctive task environments demonstrate the generalizability of the model. Despite the differences of these tasks, they share some common features. First these are human supervisory control tasks. In such tasks, automation is used more as a tool rather than a collaborative teammate. Second, the task involves infrequent human interaction and relatively high level of automation, which creates a low task load environment. Third, these are long duration tasks that require sustained attention.

In addition to the tasks used in this research, there are other tasks that share similar characteristics, such as air traffic controllers when there is light traffic, truck and rail drivers, and process control operators supervising plants operating in an automated mode. Autonomous cars are in this category as well. Autonomous cars often require the driver to supervise the system. This seemingly very easy task is actually quite difficult. Humans cannot maintain their attention and vigilance over long periods of time. Boredom, distraction, and failure in responding to emergency can and will occur in autonomous driving (Cummings and Ryan 2013). The attention degradation and performance impact could also be evaluated using the PAL model.

The model also has a few limitations. First, the PAL model made a number of simplifications and assumptions about system design, interactions, and human perception, cognition, and decision-making. The interactions with the system are simplified and represented by interaction time duration and frequency without considering lower level details about the operations. Moreover, human attention was simplified to attention on the task and attention on other activities. While these are sufficient to capture the overall trends, PAL does not fully capture the multi-faceted nature of attention. For example, visual and auditory attention are not reflected in the PAL model. Posner and Petersen (1990) proposed that attention has three separate networks: an executive control system, an attention orienting system, and an attention alerting system. Consist with this framework, Washburn et al. (2015) summarized the multi-facets of attention as an attention-focusing factor that captures the intensity dimension of attention, a sustained-attention factor, and an attention-scanning factor. In the PAL model, all these three factors were reflected but not in detail. Small variations of attention were ignored.

The second limitation is the requirement for previous data. While the structure of the model is generic, previous data was required to calibrate the model to a specific task scenario. If properly calibrated, the model could make reasonably accurate predictions for evolutionary changes to existing systems. Without prior data, absolute values of the prediction would not be informative. Relative comparisons of alternative designs are still possible but must be interpreted very carefully. The PAL model could still provide insights for futuristic systems for which little data was available. Model structures regarding human behavior, such as attention and executive control resources, can be kept the same. These are fundamental relations that are unlikely to change with new system design. However, modification of the model structure might be necessary to account for the new system architecture.

The third limitation is that the PAL model is deterministic instead of stochastic. In the real world, the time required to perform the same task by the same person would not be exactly the same every time. In the PAL model, constant values of task processing rate under normal conditions are used. For example, the actual task processing rate may be influenced by feedback loops in the model. However, if all the other variables are kept unchanged, task processing rate will not change. Other flow rates in system dynamic models are of the same nature. Discrete event simulation, another commonly used simulation method, takes a different approach. In a discrete event simulation model, variables like task processing time are modeled as a probability distribution. Each time a task arrives, a random value will be drawn from the distribution. This means the task processing time is likely to be different even throughout the same simulation when all the other variables are kept the same. Although this limitation did not significantly impact the prediction on performance in this research, the range of variation of attention might be better captured using a stochastic approach. Moreover, a deterministic model yields a point value for a given set of parameters comparing to a range out outcomes as produced by stochastic models. In some cases, stochastic and deterministic models could lead to different choices of system improvement policies (Rahmandad and Sterman 2008). This limitation could be improved by including stochastic approaches in addition to typical SD modeling. For example, random distributions can be used to replace constant values.

Fourth, the PAL model generates average level of attention and performance of a sample population while ignoring the heterogeneity of individuals. One example of this is reflected in the discrepancy between model prediction and experiment data on performance presented in Chapter 5. The model used the average task processing time of the participants, while it was a bimodal distribution in the experiment data. There are other average values used in the model, such as personal precursors, time to distract, initial attention level, etc. While individuals always differ from each other, an important tradeoff in system dynamics modeling is between predicting the general behavior changes and capturing individual variability.

Another limitation of the PAL model is the high level of sensitivity exhibited when people have low level of engagement. Sensitivity analyses on model parameters were described in Appendix E. The most sensitive parameters identified are normal processing rate, and those related to active fatigue, passive fatigue and effect of stress on attention. While parameters were expected to have an impact on model output, the high level of sensitivity calls for improvement for the model structure related to the nonlinear relationships and more effort for parameter estimation.

In addition to the limitations of the PAL model, the data used in this study also have limitations. Most of the participants of the experiments were students or people with no experience with safety critical systems. This had two main impacts on their performance: expertise and motivation. The impact of limited experience and training could result in longer task completion times, higher error rates, and larger variance among participants for performance. When calibrating the model with such data set, it represents behaviors and performance of novices instead of experts. Moreover, since simulation platforms were used in the experiments, the participants were likely to be less motivated than real operators. Failing to clear the alarm in nuclear power plant or track all threatening objects in the real world would lead to disastrous consequences, and consequently be more stressful environments. However, these factors could not be adequately represented in the simulation platforms. While these are common issues for lab experiments, we must carefully examine their impacts, especially if we want to use the PAL model in the real world. In Appendix G, the impact of expertise in PAL is examined using Monte Carlo simulation. It shows that experts, who are faster and more consistent in task processing, have better performance and less variation.

7.3 Model Applications

The model can be used to understand the change of attention in low task load automated environment and the impact of such changes on task performance. More importantly, PAL can be used by system designers to test the effectiveness of different interventions in helping operators maintain attention and improve performance.

First, designers can use the PAL model to explore the effect of system changes on attention and performance. In Chapter 4 (Section 4.4), the effect of attention alert was predicted and compared with experiment data. It was shown that attention alerts could increase overall attention level and improve performance only slightly. In Chapter 5 (Section 5.4), the effect of restricting external distraction sources was predicted and compared with experiment data. It was shown that performance was greatly improved when external distraction sources were eliminated. In addition to these two designs, another system improvement approach, named 'increase task engagement', was proposed in Chapter 4 (Section 4.5), and the impacts on attention and performance were predicted using the PAL model. If distraction and mind wandering could be directed to increasing task engagement (more self-imposed events), attention management and performance can be improved. Designers could also assess the effectiveness of other system changes quickly using the PAL model by carefully incorporating their impact on attention related variables in the model.

Second, the PAL model can be used to explore the impact of personal precursors. As shown in Chapter 6, personal precursors, such as boredom proneness and sleep, influence the level of attention and the performance in responding to an emergency event. By varying the value of these two variables, performance can be either improved or decreased. In Chapter 6 (Section 6.5), two extreme values of personal precursors were tested using the model. This could facilitate setting the thresholds for personnel selection. Other personal precursors may also have impacts in low task load environments. They can be added in the model given that support from empirical results is established.

Third, the PAL model can be used to explore design space on automation characteristics and task load requirement. The increased level of automation creates low task load environments as investigated in this research. They are characterized by low frequency of human interactions that can be completed in short time under normal circumstances. Although the task load can be high when an emergency event happens, the long duration before the onset of an emergency event often results in boredom, frustration and distraction. In the PAL model, the impact of high level of automation is reflected in the task arrival process and normal task processing rate. By changing these variables, designers could use the PAL model to assess the change in level of automation on attention and performance. For example, will the performance be improved if the operator needs to do some simple tasks when there is no emergency event? This question was addressed in Chapter 5 (Section 5.5). What if the task processing rate is changed as a result of added tasks or improved interface design? This question was addressed in Chapter 6 (Section 6.5):

- *Adding a Testing Task* (Section 5.5). If a system testing task was processed during the low task load period prior to emergency event onset, attention management and performance can be improved. The improvement can be even larger if the testing task results in familiarity with the system and better situation awareness.
- *Changing Task Processing Capability* (Section 6.5). If task processing capability is increased, attention management and performance can be improved. The

improvement on performance is affected by the attention and ECR level, which is influenced by event onset times and individual differences.

Testing these approaches will not directly generate a detailed system design. However, they could provide guidance for designers.

7.4 Contributions

The objectives of this research are to systematically investigate the attention and performance of human operators when they interact with automated systems under low task load, and build a dynamic model and use it to facilitate system design. In the process of achieving these objectives, several contributions were made to the domain of human automation interaction in low task load environments, including both theoretical, modeling and applications.

In Chapter 1, the first research question asks "What are the major factors and influences that affect boredom of human operators when they interact with automated systems? Which of these factors and influences should be captured in a model?" In order to answer this question, a thorough literature review was conducted, which led to a systematic framework called the Boredom Influence Diagram. It is the first such systems representation of boredom and its multidimensional attributes. This framework highlights the distinction between low task load environments resulting from high levels of automation and the repetitive tasks in manufacturing and traditional vigilance tasks. Building on this framework, five interconnected components were represented in a dynamic model using system dynamics modeling method.

The second contribution is a dynamic model of human attention and performance in low task load automated environments. While there are few studies investigating low task loading in automated environments, models of such systems are scarce. The PAL model is a novel application of the system dynamic modeling method in human automation interaction. Moreover, PAL makes it possible to observe the changes in attention and performance over time rather than average values or measurements at only a few discrete time points. Three experimental datasets were used to assess the validity of the model in capturing and predicting human attention and performance in three different low task load settings. This answers the second and third research question presented in Chapter 1 regarding the prediction ability and level of accuracy of the model.

Practically, system designers can use the PAL model to assess the impact of design changes and personal precursors in order to help maintain attention level and improve performance, as discussed in the previous section. This allows fast exploration of design alternatives with low cost.

7.5 Future Work

Given that automation is becoming more prevalent in complex and simple systems, more research is needed in mitigating negative consequences as a result of long periods of inactivity and boredom, for both experts who are highly trained and for widely varying populations such as those in driving domains. Three main areas of future work have been identified: 1) Further investigation of individual differences to facilitate personnel selection, 2) Development of system and task designs to mitigate the negative consequences, and 3) Improve and extend the PAL model.

7.5.1 Individual Differences and Personnel Selection

While boredom proneness and sleep quality have been investigated in this research, there are more factors that could influence attention management and performance in low task load environments. Moreover, low task and high task load within the same mission may be influenced by different characteristics of individuals. Such interactions of the environment with individual differences have not been studied in any depth. In one air traffic control task study, it has been suggested that task characteristics of repetitiveness and traffic density may interact with individual influence (e.g. personality, experience, age) in a way that causes monotony and boredom (Straussberger and Schaefer 2007), but more work is needed in this area. With distractions such as smart phones so readily available, the link between perceived boredom and distraction is another area that deserves more focus. Washburn et al. (2015) suggested that selection, training, and assignment of individuals in applied-perception contexts should be guided by individual differences in the capacity to maintain executive attention in the face of competing experiential and environmental constraints.

Although it will not be possible to preferentially select operators who are not prone to boredom in domains where automated technology is ubiquitous, such as driving, for other safety-critical domains like nuclear power plant control and UAV operations, screening personnel is already part of the culture. While it is unlikely that any single variable could successfully predict performance alone, some attempts have been made to evaluate personnel using a multivariate approach (Matthews Warm Shaw et al. 2010; Matthews et al. 2014). Results show that individual ability in reasoning and vocabulary, performance on short vigilance tasks, task engagement, task-focus coping, and avoidance coping explains 30% of the variance on vigilance performance of long durations. Clearly more objective quantitative data are needed in these areas to understand the interaction of the individual with tedious supervisory control environments, particularly in domains that could require time-pressured responses like those in military command and control environments, as well as process control settings.

7.5.2 System and Task Design

Two system design changes have been tested using the PAL model. Based on the strategies people use to cope with boredom as illustrated in Chapter 2, there are more design options for systems and tasks that can make the environment less boring and potentially distraction-inducing.

One basic strategy is to schedule tasks so that human operators get enough breaks and rests to recover from boredom (Azizi et al. 2010). In line with the boredom coping strategies, task design can be improved by including a secondary task that is stimulating, but not demanding. In one study, drivers that made fewer errors/misses during a monotonous laboratory task tended to experience larger variance in actual engine speed control with fewer accidents in their driving record. These drivers tended to introduce various task unrelated activities during monotonous driving such as looking for deer on the side of the road, which both reduced boredom and increased alertness (McBain 1970). Another study demonstrated that an interactive cognitive task could combat fatigue in monotonous driving environments (Gershon et al. 2009).

Additional past research has shown that monitoring performance can be improved through dividing attention across tasks (Gould and Schaffer 1967; Tyler and Halcomb 1974).

However, this research in a naturalistic setting demonstrated that operators were far less likely to divide their attention than to be completely distracted. Indeed, there is an increasing body of literature that shows that people are not as effective at dividing their attention as they might think (Loukopoulos et al. 2012; Ophir et al. 2009).

The third aspect of task design is to reconsider the level of automation. While increasing the level of automation is the goal of many system designers, they should also consider its impact when there are humans in the loop (Parasuraman et al. 2000). In a recent study on automated driving, responses to critical incidents involving an obstruction in the driver's lane were worse when distracted under automated driving conditions as compared to manual driving (Merat et al. 2012). From this perspective, decreasing the level of automation, at least partially, may be beneficial for system performance.

Two approaches that have been shown to improve performance and maintain situation awareness when monitoring automated systems are: 1) intermediate levels of automation to maintain engagement in complex system control, and 2) adaptive automation for managing operator workload through dynamic task allocation between the human and machine (Kaber and Endsley 2004). In a study looking at automation monitoring during multitask flight simulations, performance on automation failure detection was better with adaptive task allocation that temporarily returned the control from the automation to a human operator than when under full automation control (Parasuraman et al. 1996). When implementing dynamic control allocation, issues such as the decision authority, triggers of control changes, and task characteristics must be carefully evaluated (Johnson et al. 2014).

These system and task designs require more empirical work in the future. It is also hoped that the PAL model could facilitate the exploration and assessment of such design changes, although improvements and extensions of the current model are required.

7.5.3 Model Improvement and Extension

One aspect of improvement is to address the deterministic and homogenous representation inherited in system dynamics models. The model can be changed to a stochastic model by replacing the average values in task processing rate and personal precursors with probability distributions. Further modification of the model structure may be needed to capture the individual differences and the causes for such heterogeneity. For example, the bimodal distribution of task processing time in Chapter 5 may be connected with differences in cognitive abilities, familiarity with the task, or motivation. To justify these model changes, more empirical data is required.

Table 7-2: Model Extensions

Category	Constructs or Processes
Task Characteristics	Novelty of the task Signal discriminability Signal modality: visual vs. auditory Signal complexity: single cue or analysis based on multiple cues Dual tasking Memory load
Cognitive Processes	Working memory Situation awareness Learning effect
Individual differences	Cognitive abilities Experience or expertise Incentives and motivation Appraisal of personal significance of task

The model simplified the representation of automation and attention process. This can be improved by adding additional variables and feedback loops to the model. The key components that can be added are summarized in Table 7-2. It includes three categories: task characteristics, information processing, and individual differences. While these extensions might improve the model, the tradeoff between complexity and usefulness of the model must be carefully evaluated. More complex model also requires more data for calibration, which may be difficult to measure. In addition, complex models may not always generate more accurate predictions.

The presence of automation in the workplace is only going to increase, bringing a myriad of problems resulting from low task loading. For example, the mining industry is quickly moving towards almost complete automation, where minerals are automatically extracted, and then transported via automated rail to shipping hubs (Kara 2013). Driverless cars, while now in the experimental stage, are optimistically projected to be available to the general public by 2020 (Gannes 2014). While the automated advances in these systems could increase safety and efficiency, these and other such supervisory control systems will require a human to at least be in the loop, and able to intervene when systems degrade or fail. However, these same systems will likely induce boredom and difficulty in maintaining attention when they reliably operate for long periods of time, and how to design the system, including appropriate function allocation, will be critical.

Appendix A. Nonlinear Relationships in the PAL Model

This section visualized the nonlinear relationships used in the PAL Model. The equations for these relationships were described in Section 3.3.

Corresponding to OPS-USERS Task in Chapter 4

The positive and negative effects of stress were described in Equation (6) and (7). The corresponding parameters used for k_1, k_2, c_1 and c_2 to generate Figure A-1 were presented in Table 4-2. The active and passive fatigue were described in Equation (13) and (14). The corresponding parameters used for k_5, c_5, m_5, k_6 and c_6 to generate Figure A-2 were presented in Table 4-2.

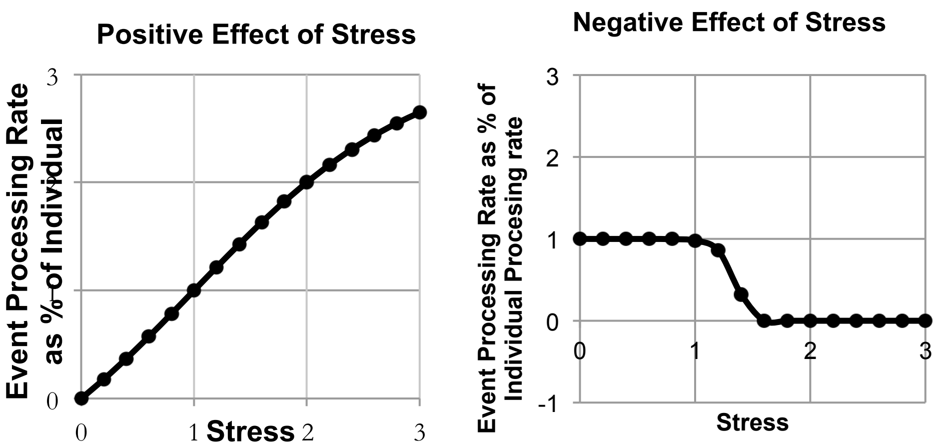


Figure A-1: Positive and Negative Effect of Stress

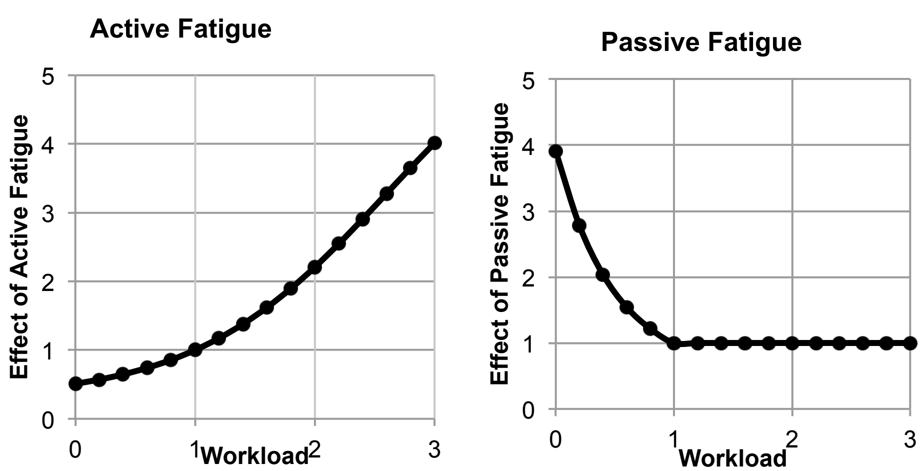


Figure A-2: Active and Passive Fatigue

The effect of stress on attention was described in Equation (20). The corresponding parameters used for k_7, and c_7 to generate Figure A-3 were presented in Table 4-2. The effect of vigilance was described in Equation (25) and the effect of ECR was described in Equation (17). The corresponding parameters used for k_3 and k_4 to generate Figure A-4 were presented in Table 4-2.

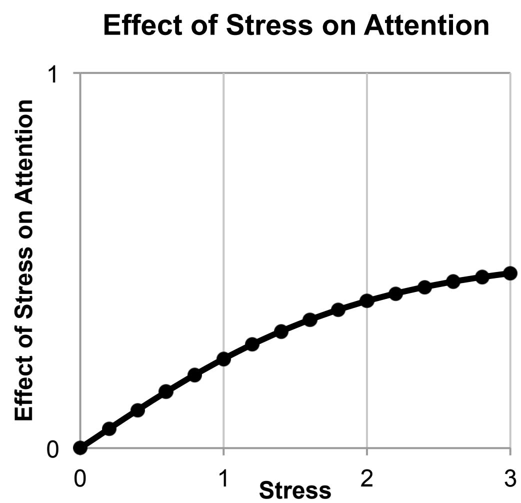


Figure A-3: Effect of Stress on Attention

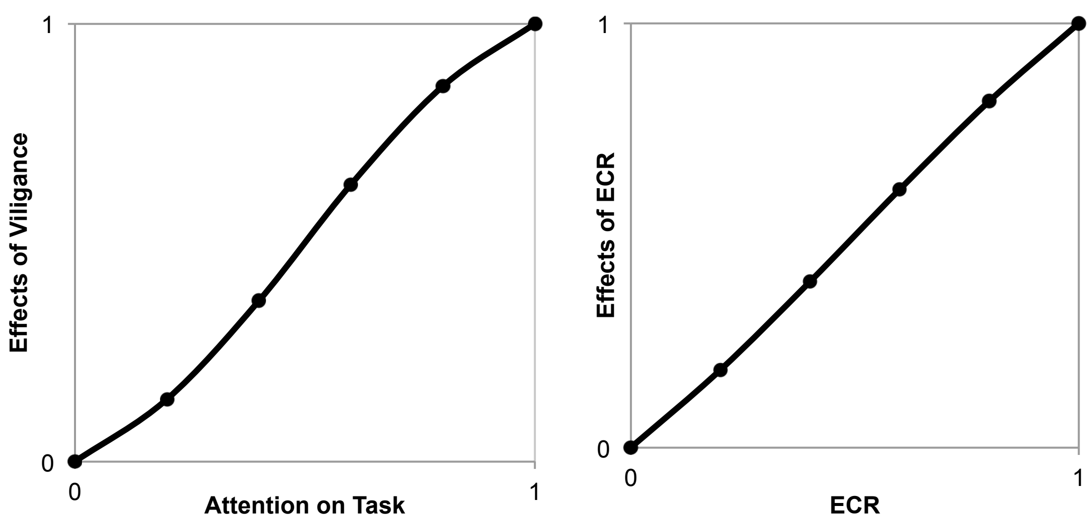


Figure A-4: Effect of Vigilance and ECR on Individual Processing Rate

Corresponding to the Nuclear Power Plant Monitoring Task in Chapter 5

The positive and negative effects of stress were described in Equations (6) and (7). The corresponding parameters used for k_1, k_2, c_1 and c_2 to generate Figure A-5 were presented in Table 5-4. The active and passive fatigue were described in Equation (13) and (14). The corresponding parameters used for k_5, c_5, m_5, k_6 and c_6 to generate Figure A-6 were presented in Table 5-4.

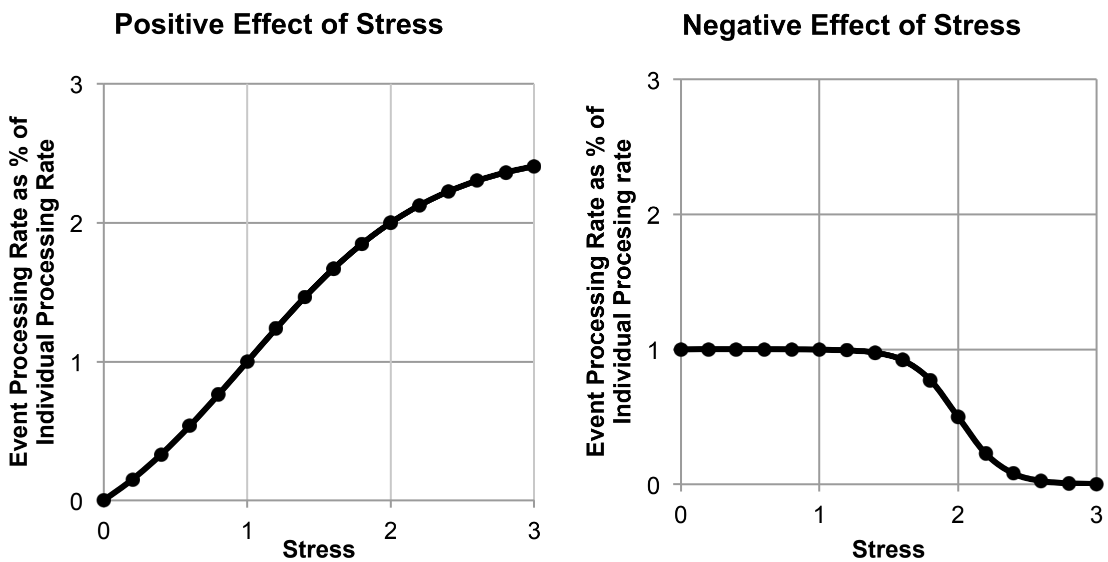


Figure A-5: Positive and Negative Effect of Stress

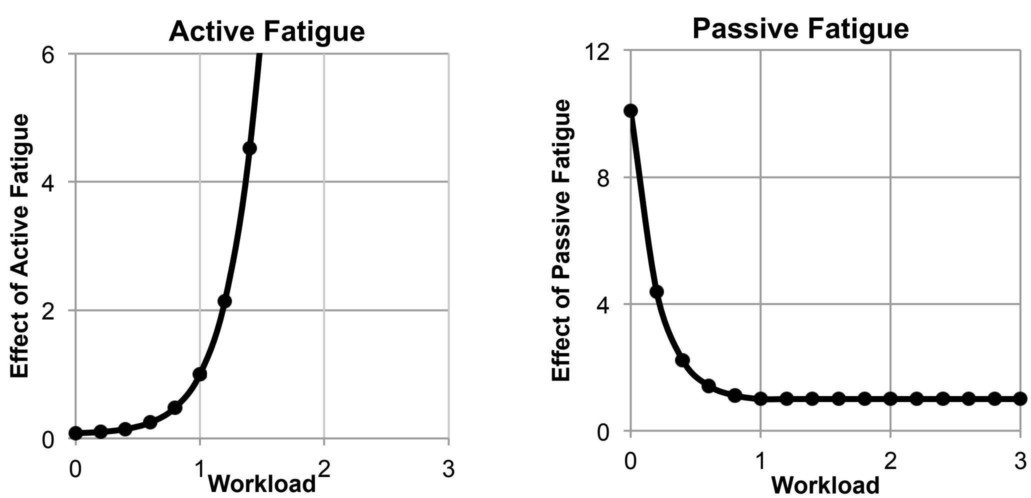


Figure A-6: Active and Passive Fatigue

The effect of stress on attention was described in Equation (20). The corresponding parameters used for k_7, and c_7 to generate Figure A-7 were presented in Table 5-4. The effect of vigilance was described in Equation (25) and the effect of ECR was described in Equation (17). The corresponding parameters used for k_3 and k_4 to generate Figure A-8 were presented in Table 5-4.

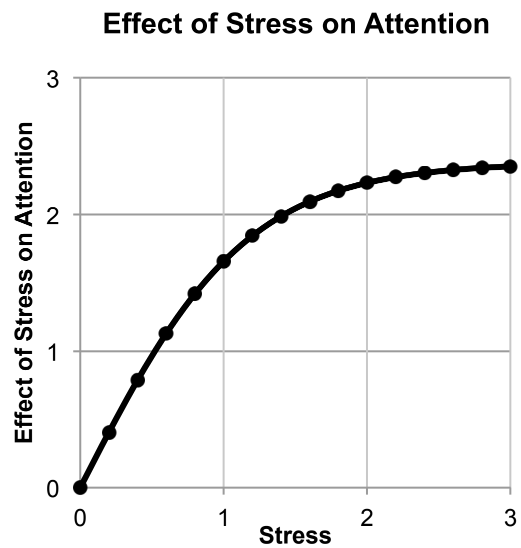


Figure A-7: Effect of Stress on Attention

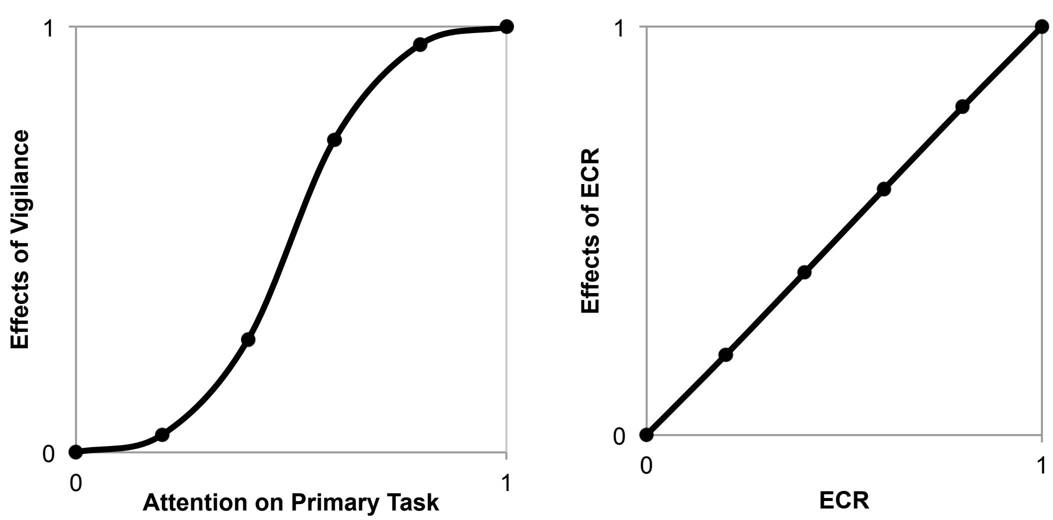


Figure A-8: Effect of Vigilance and ECR on Individual Processing Rate

Corresponding to the Tracking Task in Chapter 6

The positive and negative effects of stress were described in Equation (6) and (7). The corresponding parameters used for k_1, k_2, c_1 and c_2 to generate Figure A-9 were presented in Table 6-2. The active and passive fatigue were described in Equation (13) and (14). The corresponding parameters used for k_5, c_5, m_5, k_6 and c_6 to generate Figure A-10 were presented in Table 6-2.

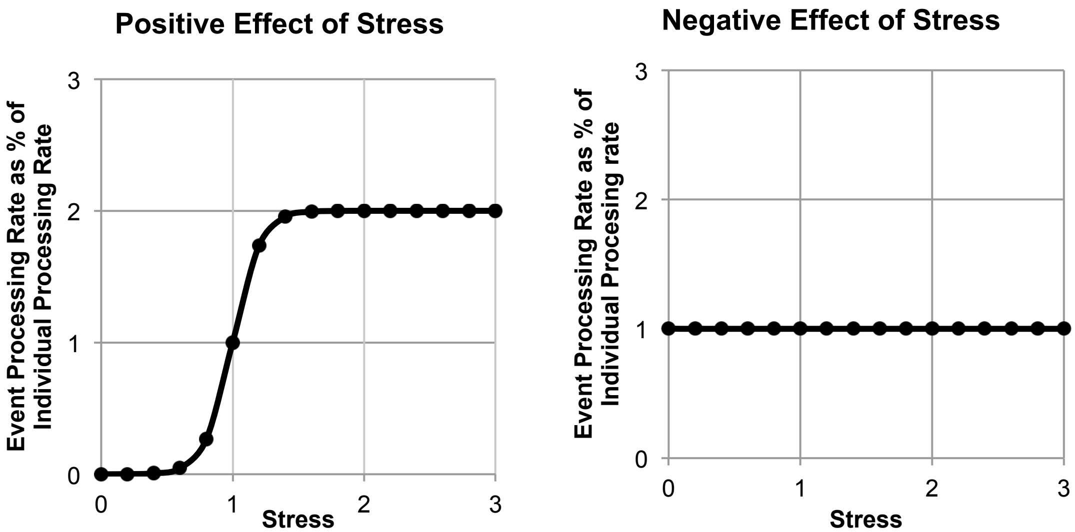


Figure A-9: Positive and Negative Effect of Stress

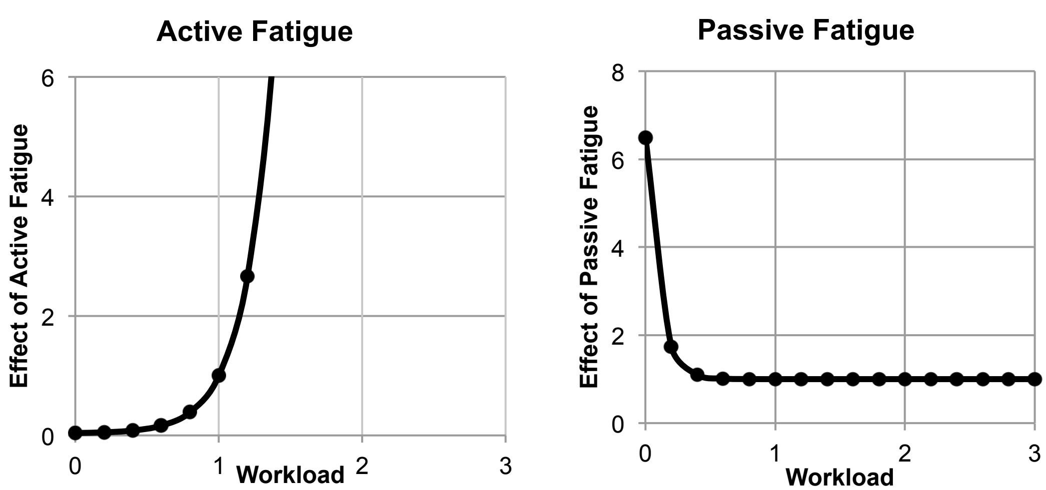


Figure A-10: Active and Passive Fatigue

The effect of stress on attention was described in Equation (20). The corresponding parameters used for k_7, and c_7 to generate Figure A-11 were presented in Table 6-2. The effect of vigilance was described in Equation (25) and the effect of ECR was described in Equation (17). The corresponding parameters used for k_3 and k_4 to generate Figure A-12 were presented in Table 6-2.

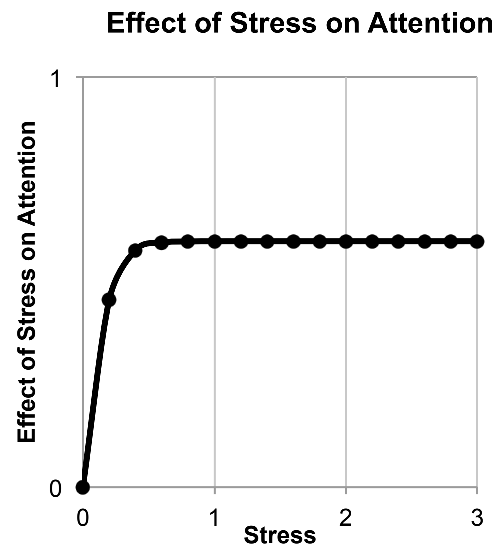


Figure A-11: Effect of Stress on Attention

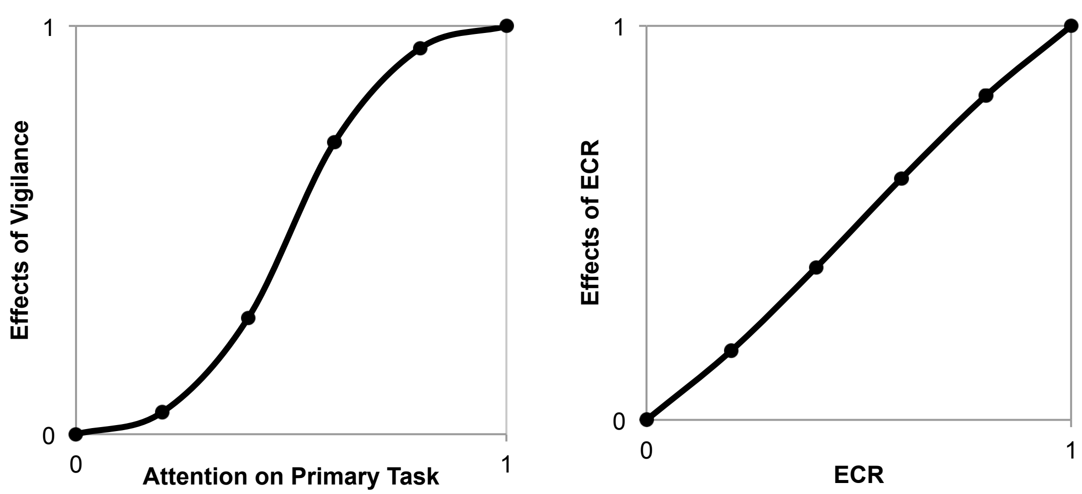


Figure A-12: Effect of Vigilance and ECR on Individual Processing Rate

Appendix B. Boredom Proneness Scale

1. It is easy for me to concentrate on my activities.	T | F
2. Frequently when I am working I find myself worrying about other things.	T | F
3. Time always seems to be passing slowly.	T | F
4. I often find myself at "loose ends," not knowing what to do.	T | F
5. I am often trapped in situations where I have to do meaningless things.	T | F
6. Having to look at someone's home movies or travel slides bores me tremendously.	T | F
7. I have projects in mind all the time, things to do.	T | F
8. I find it easy to entertain myself.	T | F
9. Many things I have to do are repetitive and monotonous.	T | F
10. It takes more stimulation to get me going than most people.	T | F
11. I get a kick out of most things I do.	T | F
12. I am seldom excited about my work.	T | F
13. In any situation I can usually find something to do or see to keep me interested.	T | F
14. Much of the time I just sit around doing nothing.	T | F
15. I am good at waiting patiently.	T | F
16. I often find myself with nothing to do-time on my hands.	T | F
17. In situations where I have to wait, such as a line or queue, I get very restless.	T | F
18. I often wake up with a new idea.	T | F
19. It would be very hard for me to find a job that is exciting enough.	T | F
20. I would like more challenging things to do in life.	T | F
21. I feel that I am working below my abilities most of the time.	T | F
22. Many people would say that I am a creative or imaginative person.	T | F
23. I have so many interests, I don't have time to do everything.	T | F
24. Among my friends, I am the one who keeps doing something the longest.	T | F

Appendix C. Model Iterations

The PAL model was developed with several iterations. With careful examination of the model variables versus previous literature, the model structure underwent significant changes. The final model (Figure 3-4) was also greatly simplified from earlier versions. Key steps of the iteration process are presented in Figure C-1 through Figure C-4. The changes can be summarized in four aspects: modeling the executive control process, simplifying the attention process, adding the Yerkes-Dodson loops, and combining boredom with passive fatigue.

Modeling Executive Control Process

In the first version of the model, executive control process was modeled as a single variable called Self Regulation, as shown in Figure C-1. It was then changed to a stock and flow process in Figure C-2, where Self Regulation Power was depleted as the boredom level increases. This variable was renamed as Mental Energy in Figure C-3, and finally Executive Control Resource (ECR) in Figure C-4. An inflow to ECR was added in Figure C-4 to capture the restoration of ECR.

This iteration process was based on literature investigation on the mechanism for decreased sustained attention on a primary task in vigilance tasks. In previous research, there are several theories explaining the phenomenon. One debate concerning these theories is the role of underload versus overload. In the underload account, the monotonous and understimulating nature of the task leads to withdrawal of attention from the primary task (Manly et al. 1999). In the overload account, it was argued information-processing abilities are depleted because vigilance tasks are taxing and effortful (Warm Parasuraman et al. 2008). While neither theory is sufficient to explain all the relationships among task demand, attention and performance, we adopted a new approach that combined both theories as proposed by Thomson et al. (2015), in which the degree of executive control decreases as time on task increases, driving the redistribution of attention. The model iterations show changes from executive control approach, to resource depletion approach, to the combined resource-control approach.

Researchers also disagrees on limited executive control resource versus motivation shifting for self-control failure (Muraven and Slessareva 2003), and whether executive control is a limited resource or can be expanded with will power (Baumeister et al. 1998). These are not addressed in the current model, since there is no consensus and the understanding of vigilance and sustained attention is still evolving.

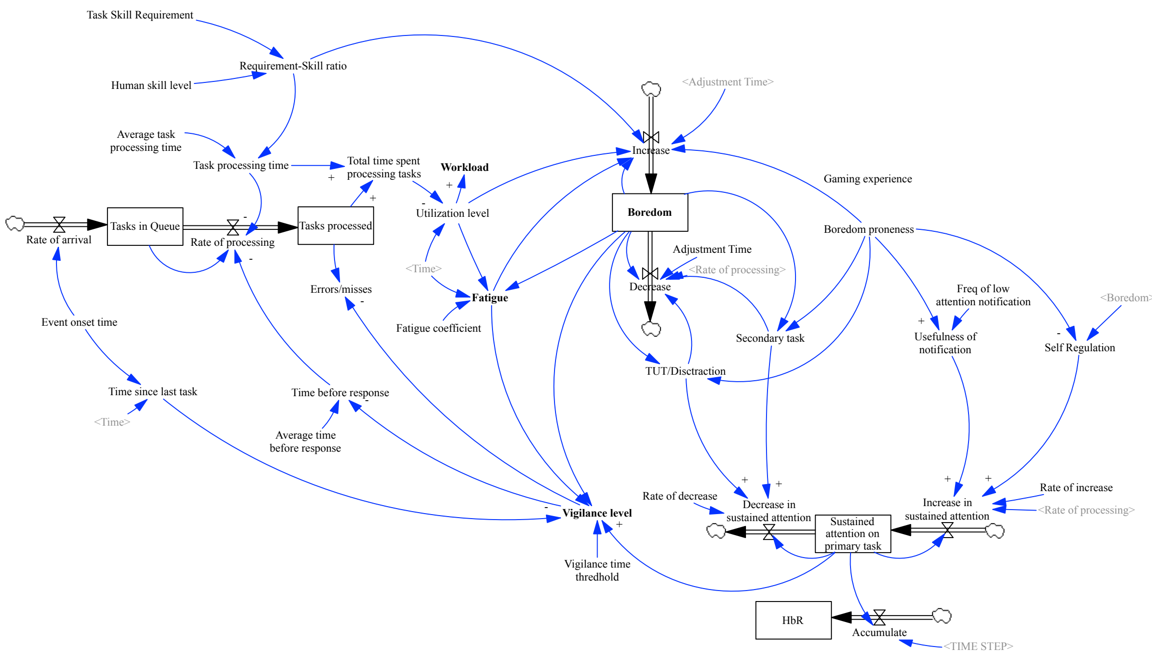


Figure C-1: Model Iteration 1

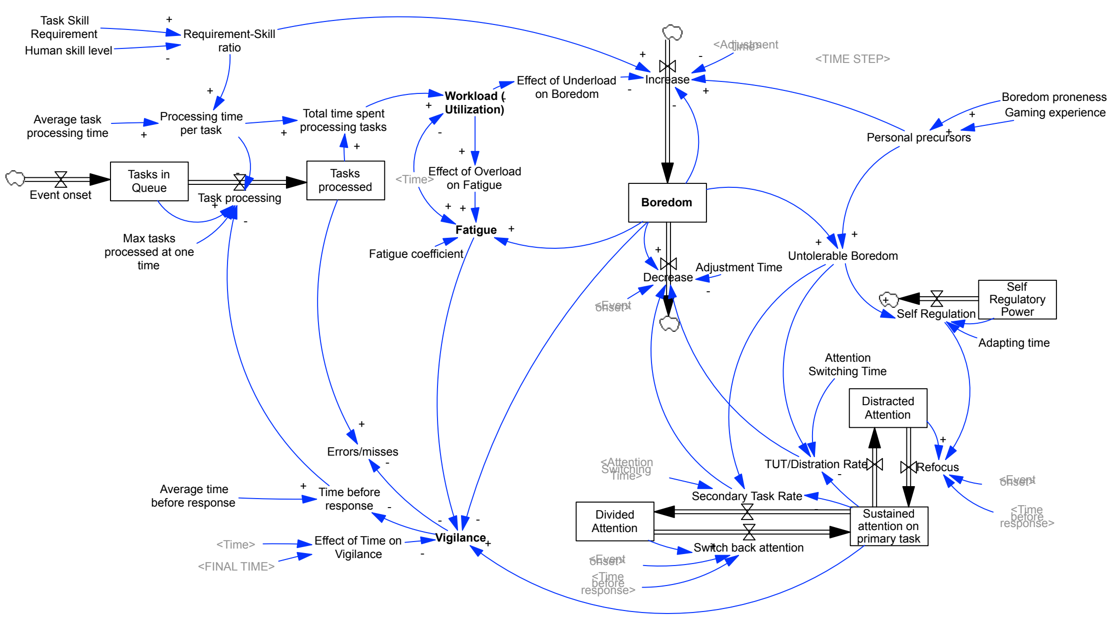


Figure C-2: Model Iteration 2

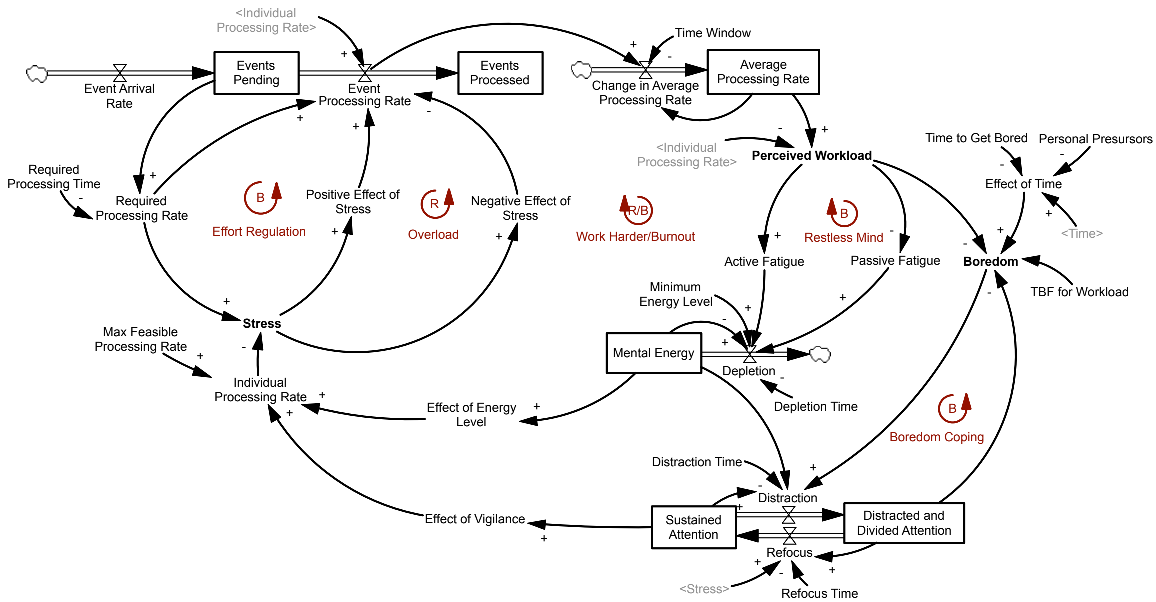


Figure C-3: Model Iteration 3

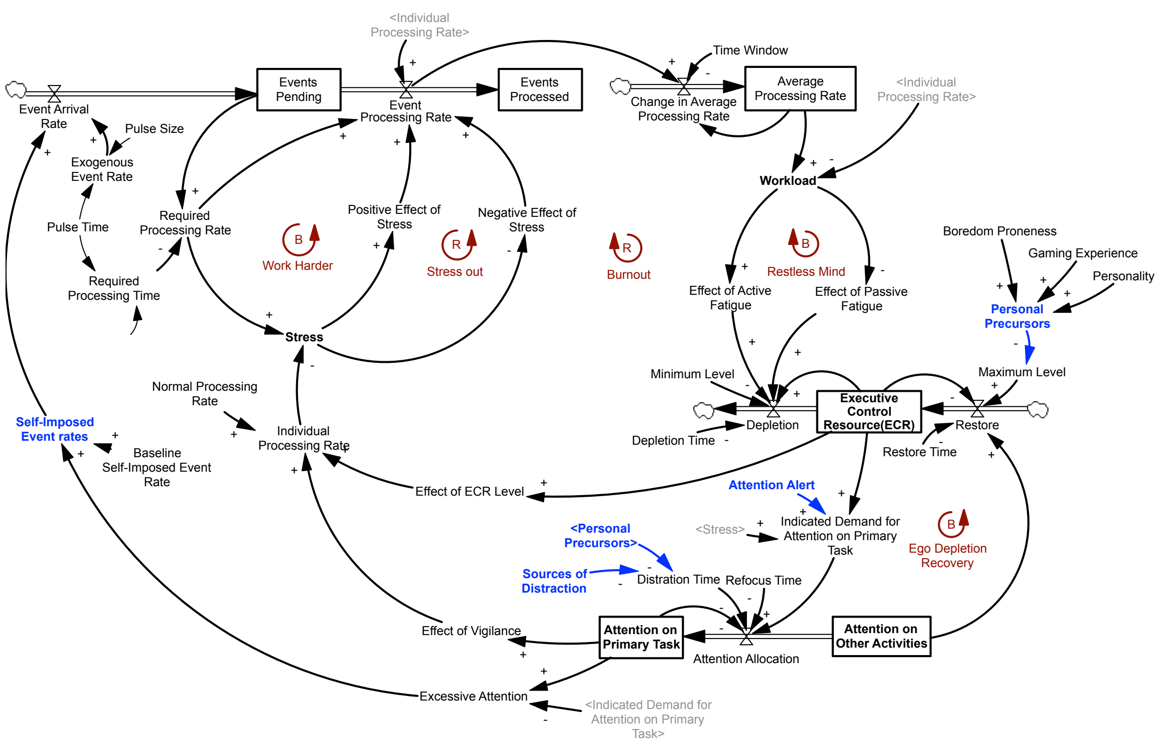


Figure C-4: Model Iteration 4

Simplifying the Attention Process

The model for attention process underwent two changes. The first was the change on number of stocks and flows, which was greatly reduced in the final model from a maximum of seven to two. This results in a simpler model, which could capture the attention change.

Attention process was modeled as a one stock two flow process in Figure C-1, a three stock four flow process in Figure C-2, a two stock two flow process in Figure C-3, and a two stock one flow process in Figure C-4. In the final model (Figure 3-4), a one stock one flow structure was used. During this process, Distracted Attention and Divided Attention were combined as Attention on other Activities because the main concern is whether people are devoting attention on the primary task or not. Distinguishing different modes of attention does not increase the quality of prediction of the model, and for the purposes of predicting task performance, was not a critical factor.

The second aspect of change in modeling attention process was using information delay in the final version instead of material delay in the earlier versions from Figure C-1 to Figure C-3. In the final model, attention was gradually adjusted to an intended level, which was determined by executive control resource and stress level. This connects the executive control process and attention process together, and is consistent with the resource-control approach we adopted.

Adding the Yerkes-Dodson Loops

The Yerkes-Dodson Law (Teigen 1994) was not included in the initial model. It was added to the model during the iteration (Figure C-3) to capture the impacts of stress on task processing and performance, which generally follows an inverted U-curve. This is essential for modeling the attention changes and performance in cases with transition between low and high task load (Chapter 5 and 6). In section 3.4.3, the impacts of the Yerkes-Dodson law were tested using a model structure test.

Combining Boredom with Passive Fatigue

The last major change was simplifying the structure for modeling boredom. Boredom was modeled as a stock in the initial version Figure C-1, and was combined with Passive Fatigue in Figure C-4. There are two reasons for this change. First, the causal relation between boredom and attention change is not clear, as empirical studies on boredom are limited. Instead of directly impacting attention distribution as a result of boredom, it is more likely that boredom changes the degree of executive control, which then drives the attention change, as described in Dynamic Hypothesis 1. Second, passive fatigue and boredom result in similar changes on attention in the scope of PAL model. Passive fatigue (Desmond and

Hancock 2001) describes the fatigue that stems from supervisory tasks, which often results in decreased task engagement and inattention. Given that passive fatigue and boredom are inextricably linked, we decided that the effects of passive fatigue and boredom could be combined to simplify the model.

In addition to these four major aspects of changes for the model, other variables and structures were also adjusted. In the final model, each of the loops was tested individually as shown in section 3.4.3.

Appendix D. Model Calibration Settings

Model calibration using Vensim® has three steps:

1) In the first step, the user selects the output variables in the model and the corresponding data series. In this study, for example, *Attention on Task* in the model was matched with the time series attention data measured in experiments. *Events Processed* or *Targets Found* was matched with the performance data in the experiments. Performance data was processed to get the time series showing the progress of task processing over time.
2) In the second step, constant parameters that need to be adjusted are added to the optimizer. The optimization interface is shown in Figure D-1. Detailed settings of the optimizer can also be changed on this interface, such as the maximum number of iterations, type of optimizer, etc. In this study, default settings were used in all modal calibration processes.

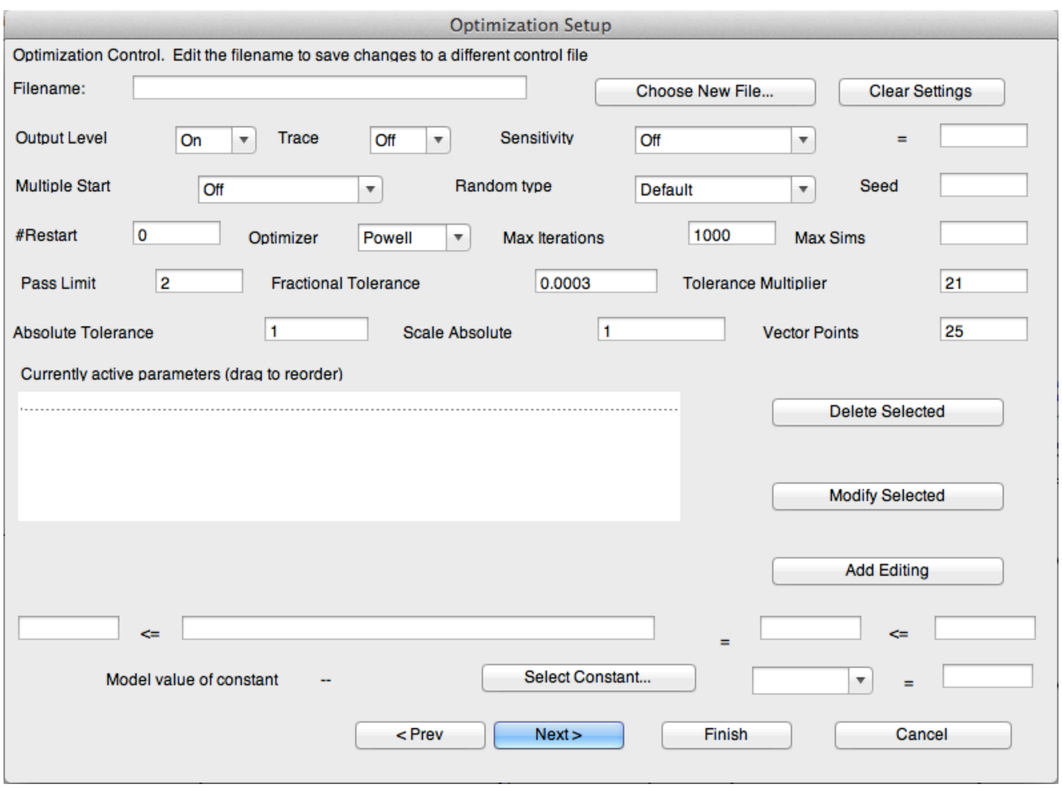


Figure D-1: Optimization Interface in Vensim®

3) In the third step, the optimizer varies the values of the selected parameters to minimize the error between model output and the corresponding data series as specified in step one. When optimization is achieved or maximum number of iterations is reached, the optimization stops.

After the calibration process, the parameter values that result in the best fit are reported. The quality of the fit as measured by MSE, R^2, U^C, U^M, U^S are reported as well. Mean absolute percentage error (MAPE) can also be calculated by the software. However, if any data value is 0 this statistic is undefined and -999 is reported for MAPE.

Appendix E. Sensitivity Analyses

Sensitivity analysis is an important test of the robustness of a computational model. It is desirable that the outputs of the model are robust to errors in parameter estimates. Sensitivity analyses help to identify the parameters for which the model outputs are most sensitive. This provides insights on the effects of uncertainties and where effort should be devoted for more accurate parameter estimations. For system dynamics models, it could also help develop system improvement policy around these parameters to change the system behavior (Eker et al. 2011). In this section, we conducted sensitivity analyses by varying parameters individually, and analyzed the impact of such changes on model outputs.

Table E-1: Parameters for Sensitivity Analysis

Task Characteristics	Individual Characteristics	Parameters for Nonlinear Relationships	
• Exogenous Event Rate	• Normal Processing Rate	k_1	Positive Effect of Stress
• Required Processing Time • Depletion Time	• Average Time to Distract • Refocus Time	k_2 c_1 c_2	Negative Effect of Stress
• Restore Time	• Time Window (for Workload)	k_3	Effect of ECR
		k_4	Effect of Vigilance
		k_5 c_5 m_5	Active Fatigue
		k_6 c_6	Passive Fatigue
		k_7 c_7	Effect of Stress on Attention

To examine the sensitivity of the model, common parameters among all three task scenarios were selected. Parameters were classified into three categories: task characteristics, individual characteristics, and parameters for nonlinear relationships (Table E-1). The first category was not used in sensitivity analysis, as we had no control over these parameters, i.e., exogenous Event Rate and Required Processing Time were determined by task requirements. Take the target tracking task described in Chapter 6 for example, Exogenous Event Rate is determined by the time the unexpected event happens. It was 40 min, 100 min or 160 min in the experiment. Required Processing Time is the time frame a task must be completed, which was determined by the experimental design. For the tracking task in Chapter 6, it was

2 min. Similarly, Depletion Time and Restore Time of Executive Control Resources were set by the length of the experiment, which can also be considered as a task requirement.

The second category includes variables related to individual characteristics. Normal Processing Rate describes how fast an average person could process a task with full attention and executive control resource[1]. Average Time to Distract and Refocus Time are time delay variables describing how long it takes to redistribute the attention. Time Window is the sampling time interval to calculate the exponential moving average of *Event Processing Rate*. It reflects the perceived workload in the recent time. In the baseline model, this variable was set to 5 minutes.

The third category includes parameters for the nonlinear relationships in the SD model. Traditionally, these nonlinear relationships between two variables are modeled using table or lookup functions. Table functions are preferred to complicated equations because the modeler can control the shape, slopes and saturation points to accurately represent the nonlinear relationship between two variables. They are also easier to interpret and visualize than complex equations (Sterman 2000). Table functions are usually constructed based on statistical studies, fieldwork, interviews, considerations of extreme conditions and physical laws (Sterman 2000). In other words, these can be quite subjective except for the case with physical laws. The use of graphical lookup tables makes it difficult to calibrate. Sensitivity analyses are rarely run for these functions (Eker et al. 2014; Eker et al. 2011). Therefore, in this study, equations are used to represent the nonlinear relationships to allow calibration and sensitivity analyses based on the parameters used in the equations.

Equations used for these nonlinear relationships are Equations (6), (7), (13), (14), (17), (20) and (25). Explanation for these equations were included in Section 3.3. These equations often described a full or a segment of S-Shaped curve, as visualized in Appendix A. They are derived based on the analytical form for a universal table function (Uys 1984). In fact, the integration of many bell-shaped probability distributions follow a S-Shaped curve. The most widely used is the cumulative distribution function (CDF) of a normal distribution, as the

[1] A Monte Carlo simulation for Normal Processing Rate is included in Appendix G, where two normal distributions were used to evaluate the impact of expertise.

central limit theorem of statistics proves the normal distribution for those accumulated events. Normal CDF links accumulations (stocks in SD models) and the probability of occurrence of related events (Franco 2007).

For the second and third categories of Table E-1, each parameter was increased by 10% and decreased by 10% in univariate testing, where parameters were varied one at a time. The impact of such changes on attention and performance at the end of the simulation are summarized in Figure E-1, Figure E-4Figure E-7Figure E-9, andFigure E-10, Figure XX, as these were the two primary dependent variables in the experiments and models. In this analysis, if the percentage of change on outputs is less than the percentage of change on the input parameter, we consider the model to be robust to the change on this parameter (Law and Kelton 2000). Otherwise, the model is sensitive to the change. Sensitivity analyses were conducted for all three task scenarios, as discussed in the following sections.

Corresponding to OPS-USERS Task in Chapter 4

The percent changes on model outputs are summarized in Figure E-1. There are a few observations. First, the model is fairly robust to changes in input parameters. All the changes on outputs were less than 10% when the input parameters were changed by 10%. Second, performance is more robust to changes in input parameters comparing to attention in general. The changes on performance are all less than 2%, while the largest change on attention is about 9%. The reason is those users are assisted by an automated planner to perform the tasks. As a result, task performance is only partially affected by human behavior changes as introduced by changing these parameters. Third, the three most sensitive parameters are m_5, which corresponds to active fatigue; c_6, which corresponds to passive fatigue, and c_7, which corresponds to the effect of stress on attention. Active fatigue and passive fatigue influence the depletion of ECR, which then affects the attention distribution. Effect of stress on attention directly affects attention when task demand imposes high level of stress. These variables are most closely related to attention in the PAL model. Hence, the output on attention is sensitive to their changes.

An additional analysis was conducted for the most sensitive parameter m_5. The impact of changing m_5 up or down by 10% on workload was presented in Figure E-2. A Monte Carlo simulation using a uniform distribution from 0.46 to 0.56 was conducted. The results for attention and performance are presented in Figure E-3. The black line shows the baseline condition, and the grey area shows the range of change. It can be seen that the range of change on attention is increasing over time with the largest deviation from baseline condition happening at the end of the task mission. This is because the effect of active fatigue on the depletion of executive control resource is accumulated over time. Another observation is that the trend of decreased attention over time is not changed. In other words, the behavior of the model stays stable with the change on m_5. For performance, the range of change is very small, which is consistent with the data shown in Figure E-1.

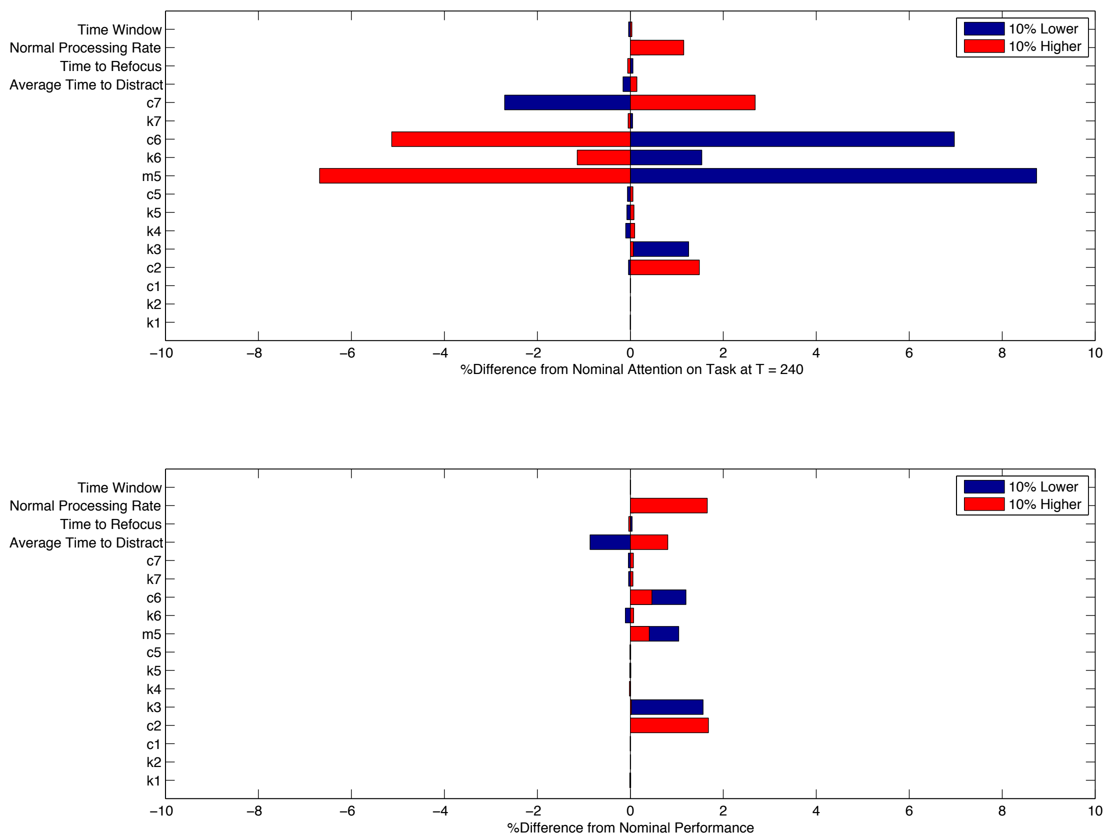


Figure E-1: Sensitivity Analysis - OPS-USERS Task

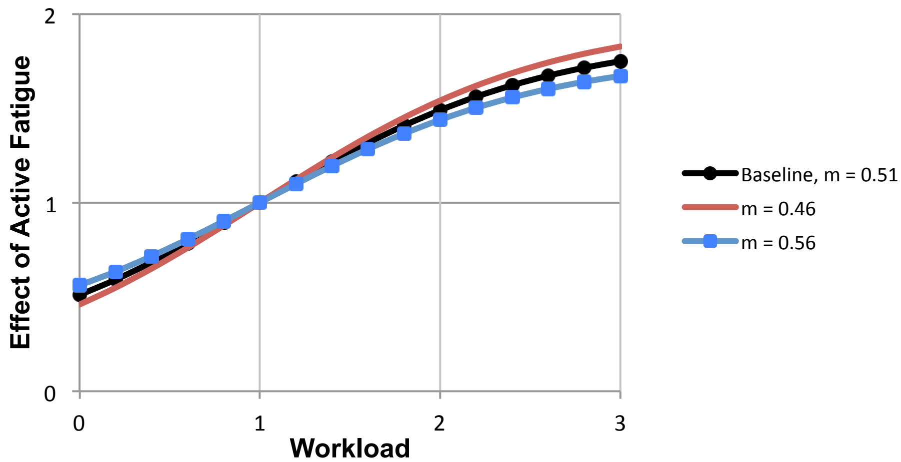


Figure E-2: The Impact of Parameter m_5 on the Effect of Active Fatigue Curve

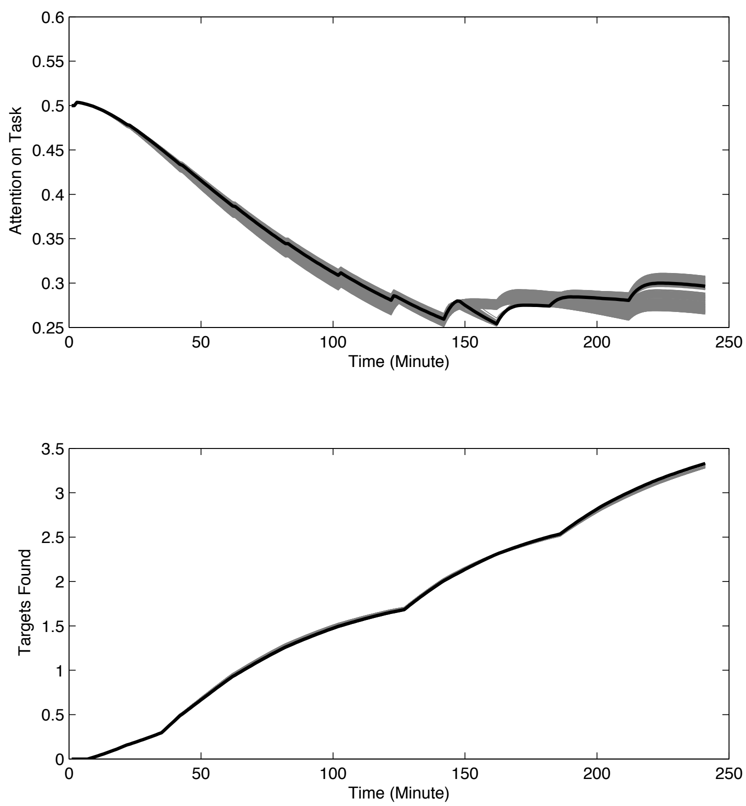


Figure E-3: The Impact of Parameter m_5 on Attention and Performance

The results in Figure E-1 also reveal asymmetries in some results, including Normal Processing Rate, c_2 and k_3. However, the percentages of changes are less than 2%. The actual change on output is also very small. Take normal processing rate for example, the baseline attention level was 0.2851, the attention level with higher normal processing rate was 0.2884, and the attention level with lower normal processing rate was 0.2856. With such small changes on outputs, the asymmetries are not big concerns for the model. The change of parameters result in a larger positive increase on outputs and almost none negative increase. The reason is that attention is bounded between zero and one. If there is no task, attention will approach zero follow an exponential decay. In other words, the decrease of attention slows down at a lower attention level. The negative increase is smaller than the positive increase because the room for change is smaller.

Some parameters did not have a large impact on attention or performance. However, these parameters are theoretically based and essential for defining the various nonlinear relationships and capturing the attention distribution process in the model. They should not be removed from the model solely based on the sensitivity analysis results.

Corresponding to Nuclear Power Plant Monitoring Task in Chapter 5

The percentages of changes on model outputs under sterile condition are summarized in Figure E-4, and those under the distraction condition in Figure E-7. There are a few observations. Under the sterile condition, early and late event onset conditions were more robust to changes in parameters. However, the middle event onset condition was more sensitive to changes in parameters. The most sensitive variables are Normal Processing Rate, c_7, k_7 and c_6. Normal Processing Rate, which was estimated based on experimental data directly relates to task performance. c_7 and k_7 correspond to effect of stress on attention, and c_6 corresponds to passive fatigue. Based on the model structure, passive fatigue and effect of stress on attention are closely related to attention distribution in the PAL model. Hence, the output on attention is sensitive to their changes. Additional analyses were conducted using Monte Carlo simulation for c_7, k_7 and c_6 to assess the level of sensitivity over the whole course of the task mission. Since system dynamics models are nonlinear, the impacts of parameter changes are likely to vary over time. A Monte Carlo simulation was conducted for c_7, k_7 and c_6 using uniform distributions ranging from -10% to 10% of their original values. Since the middle event onset – sterile condition is the most sensitive to parameter changes among the three onset times, we presented the results from Monte Carlo simulation only under this condition in Figure E-5.

The results show that the model is robust to the change of parameters before event onset and during the early stage of event processing. The variation gets larger near the end of the mission. The reason for the large variation is that people stop working on the task early in some cases, perhaps due to stress or giving up because they know the experiment is about to end. In addition, people have a lower attention and ECR level before event onset in the middle event onset condition compared to early onset. The level of attention at full engagement has a large difference from when people start to disengage with the task.

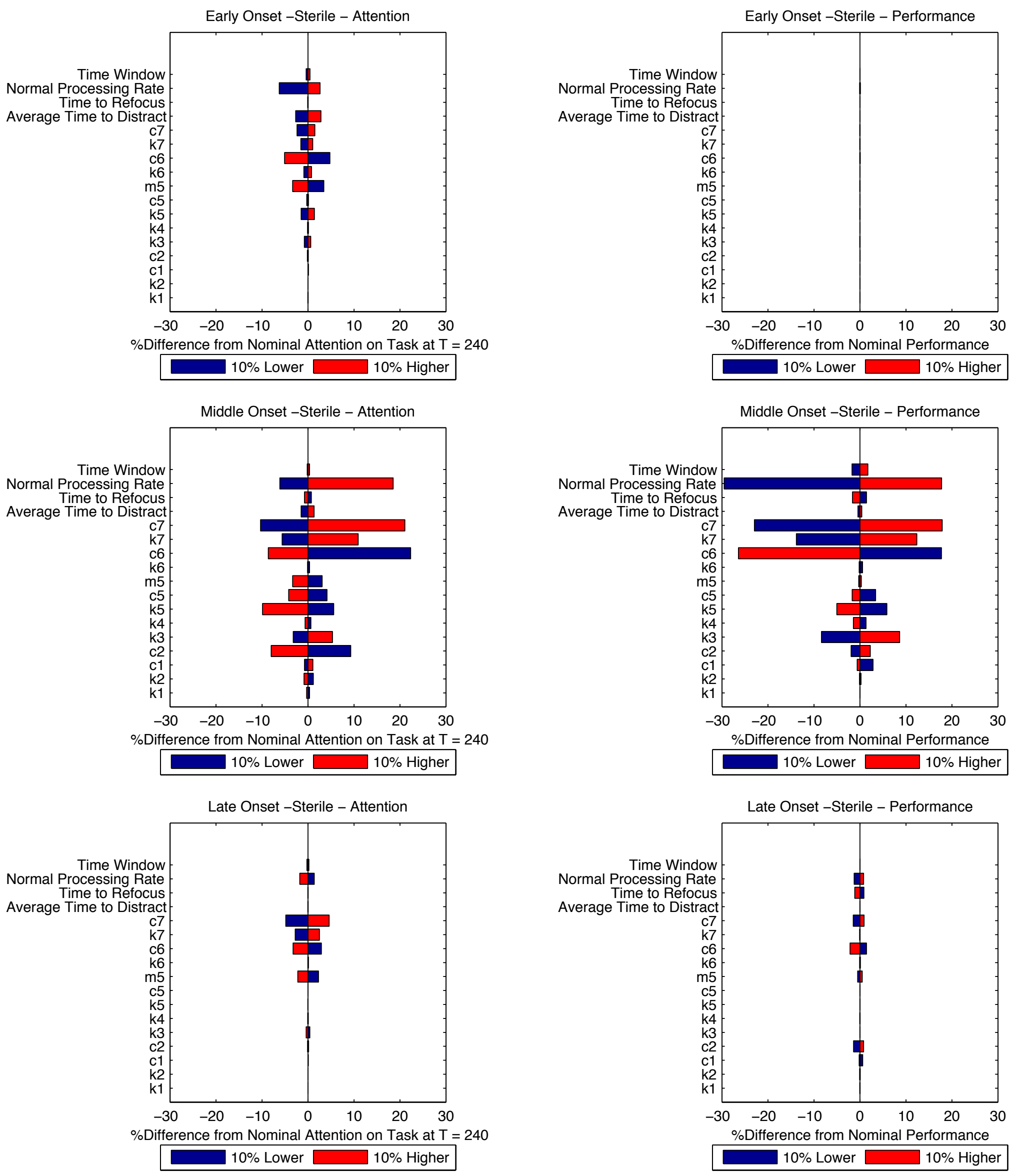


Figure E-4: Sensitivity Analysis – Sterile Condition

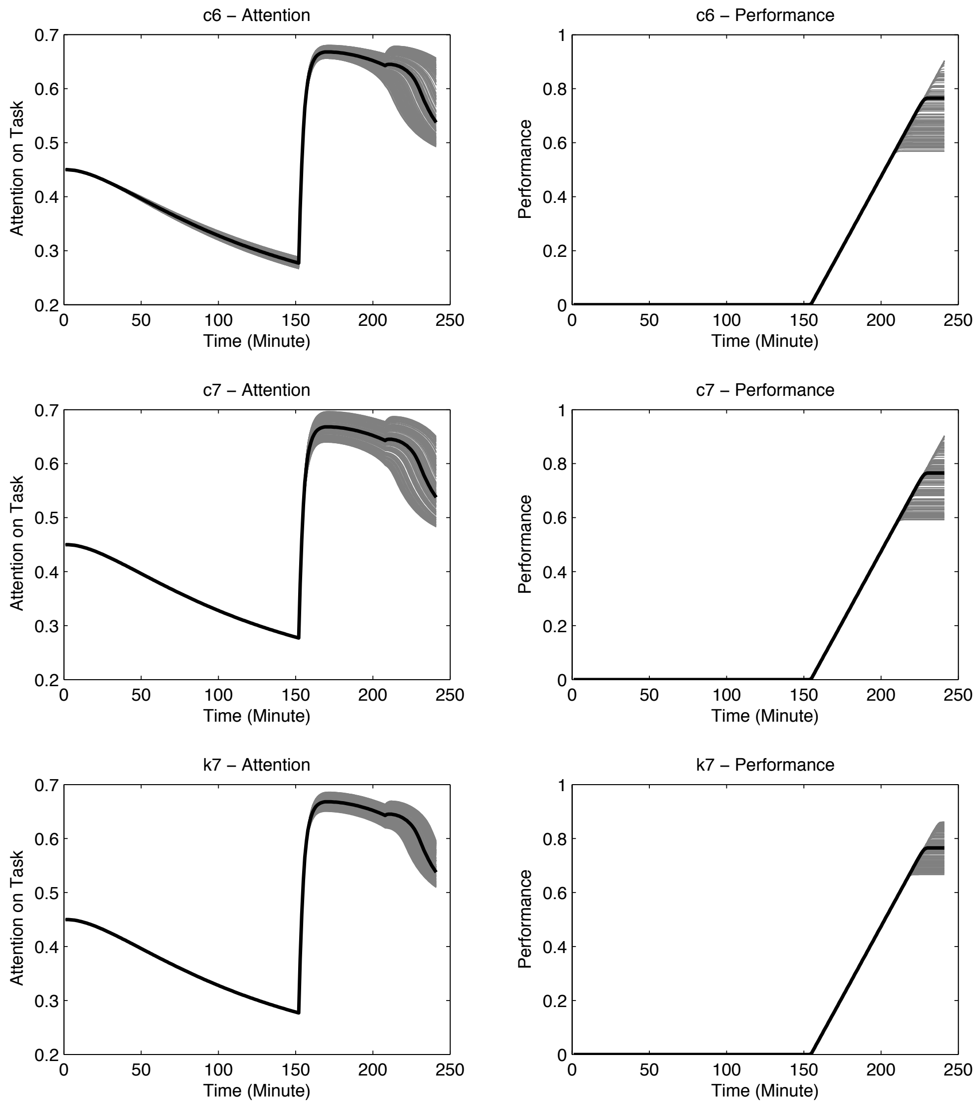


Figure E-5: Monte Carlo Simulation for Middle Onset - Sterile Condition

Such high sensitivity is not exhibited in early and late onset conditions. This can be seen clearly in Figure E-6, where the impact of c_6 under different event onset conditions was presented. In the early event onset condition, the attention and ECR level are both high. People have sufficient time to process the event and enter the distraction state gradually. It is not clear why people are in a higher attention state at the end of the experiment, but one theory is that subjects are told how long an experiment lasts and their attention and engagement may increase as they get close to the experiment conclusion. This effect has been seen in a similar study (Hart 2010). Thus our hypothesis is that people are at their lowest level of engagement in the middle of a study, and because of this difference in mental state, the model may not be accurate during the middle onset condition.

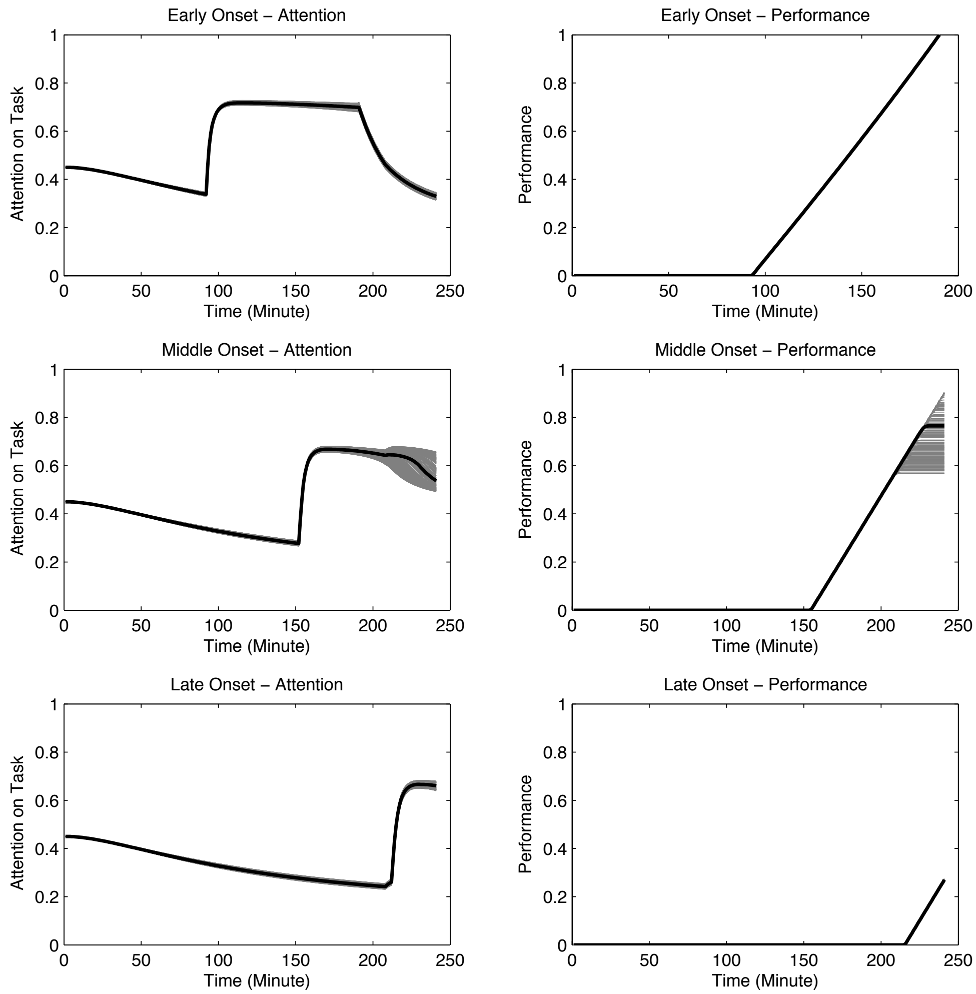


Figure E-6: Monte Carlo Simulation for c_6 under Sterile Condition

Under the distraction condition, the impacts of parameter changes were much larger than sterile condition overall, as shown in Figure E-7. It can be seen that many parameters have very large impact on model outputs. In Figure E-7, attention is represented by the value at the end of the experiment. When looking at the full range of attention change, sensitivity is not at the same level at all times. Take the impact of c_6 for example (Figure E-8), different input values result in dramatic attention changes only near the end of the task mission. This is the same as the phenomenon presented in Figure E-5, in which high sensitivity kicks in for the middle onset sterile condition near the end. In sterile condition, the level of engagement is the lowest for middle onset condition. For distraction condition, the engagement level is lower than sterile condition overall and people are more easily distracted. It is possible that the model is more sensitive to changes in parameter values when the attention level is low. This means that the model may not be well attuned to what cognitive processes are being engaged during times of distraction. In such case, more efforts should be paid to parameter estimation under such conditions.

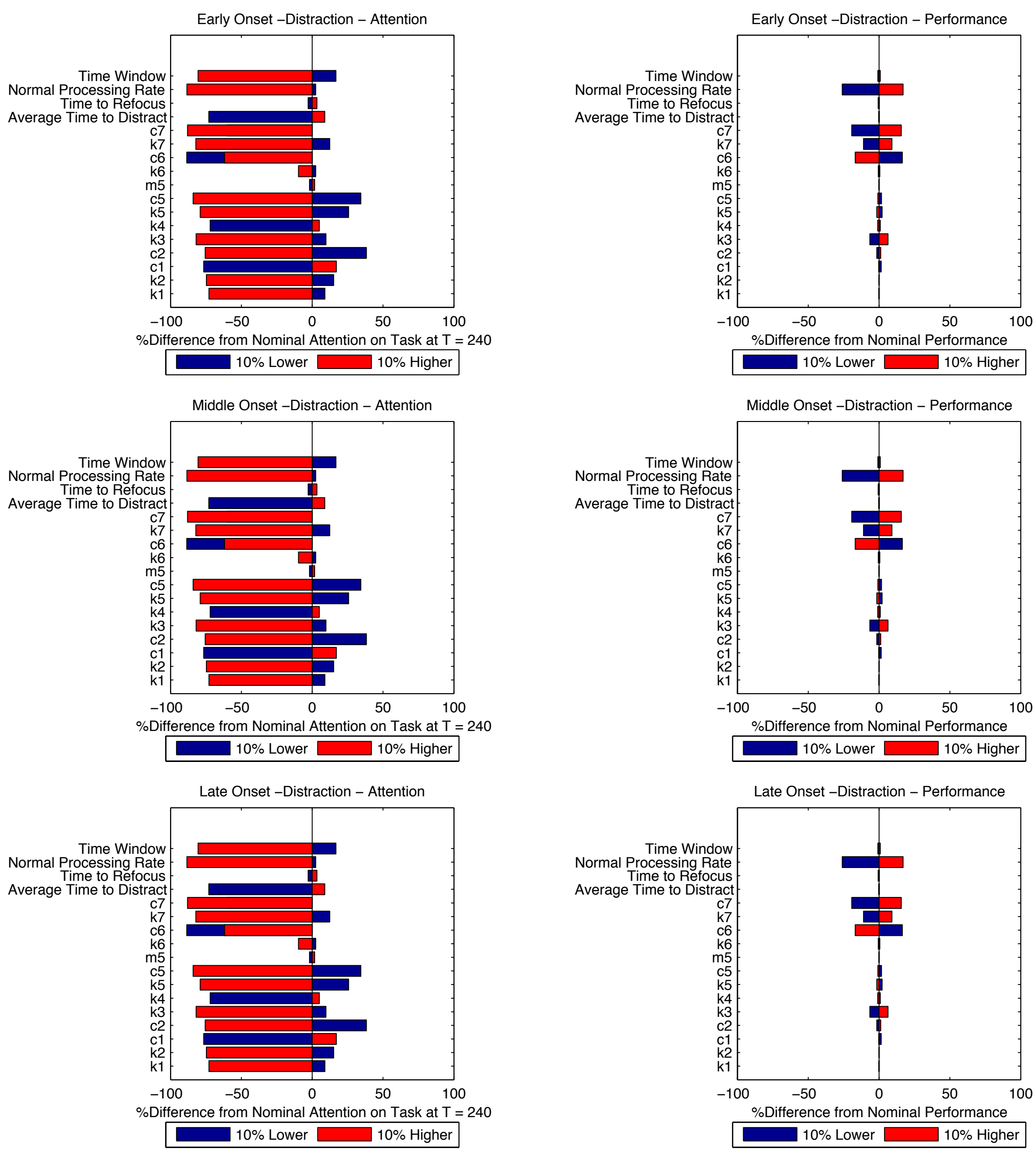


Figure E-7: Sensitivity Analysis – Distraction Condition

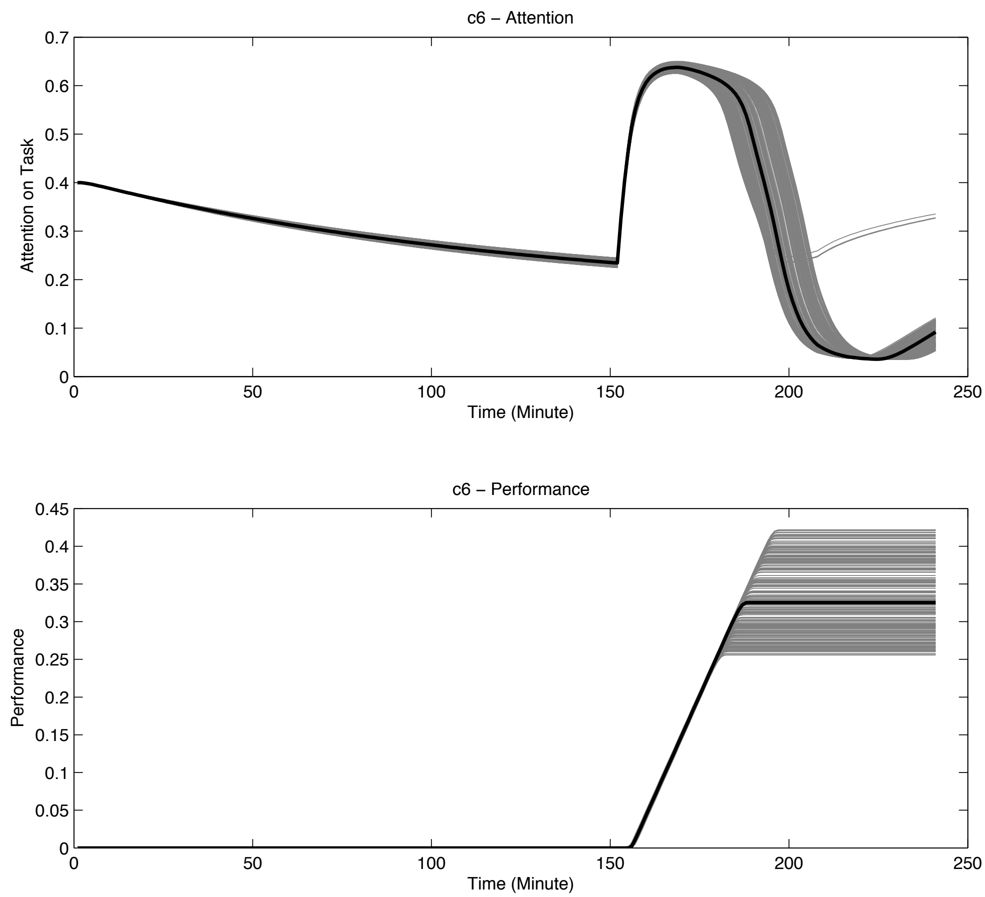


Figure E-8: Sensitivity Analysis on c_6 - Distraction Middle Onset Condition

Corresponding to the Tracking Task in Chapter 6

The percentages of changes on model outputs with the easy task are summarized in Figure E-9, and those with hard task in Figure E-10. There are a few observations. First, in general attention is robust to changes in parameter values. In this task scenario, the task processing period is only 2 minutes, which is very short comparing to the length of the task mission 180 minutes. Because of this, the impact of task processing on attention change is very small. Attention is mostly affected by the depletion of ECR when nothing happens. Second, for performance, Normal Processing Rate has a large impact. This is expected, as it directly reflects how fast the tasks are being processed.

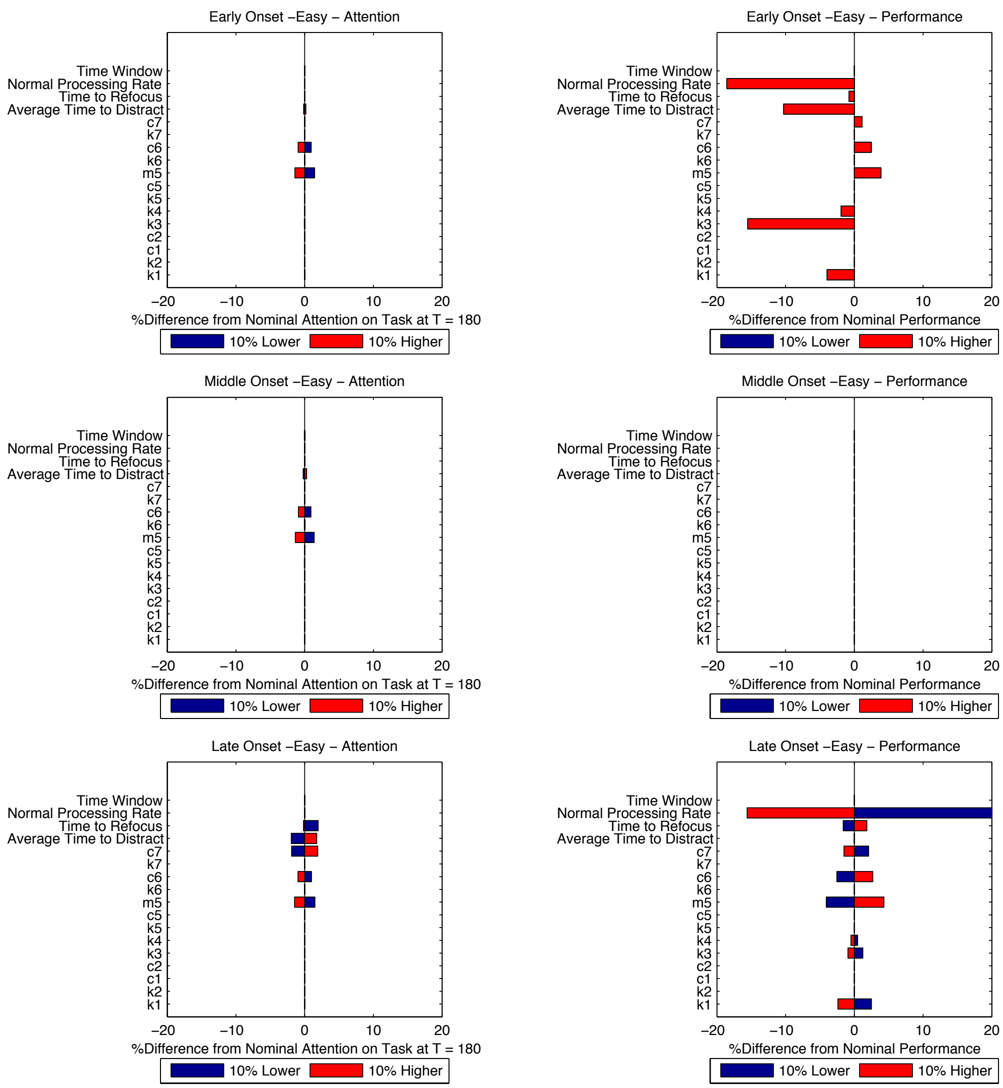


Figure E-9: Sensitivity Analysis – Easy Task

For the hard task in Figure E-9, the most sensitive parameter is Normal Processing Rate. The high sensitivity exhibits for performance in middle and late event onset conditions. In the PAL model, Normal Processing Rate directly reflects how fast the tasks are typically processed. As a result, it is expected to have a large impact on model output. However, the high level of sensitivity reveals three issues. First, the PAL model needs improvements on the part related to stress and Yerkes-Dodson loops to reduce the instability (see Figure 3-7). In the model structure, Normal Processing Rate feeds into the stress level and then affects actual task processing rate through the Yerkes-Dodson loops. A 10% increase on Normal Processing Rate causes about 10% decrease on stress. However, the Yerkes-Dodson loops amplify the changes. It is possible that an inverted U-curve with flatter top could reduce the sensitivity. In addition, instead of starting the positive effect of stress at zero (Figure A-9), it could start at a positive value smaller than one. In such case, the change on stress would cause less dramatic change on performance. Second, more efforts should be given to estimate Normal Processing Rate accurately. Currently, this variable was estimated based on experimental data. To get a more accurate estimation, operators should be given sufficient training and be measured multiple times. Third, the importance of Normal Processing Rate also means there is an opportunity for significant system improvement if we could improve it. In fact, an analysis on how Normal Processing Rate affects task performance was discussed in Section 6.6.

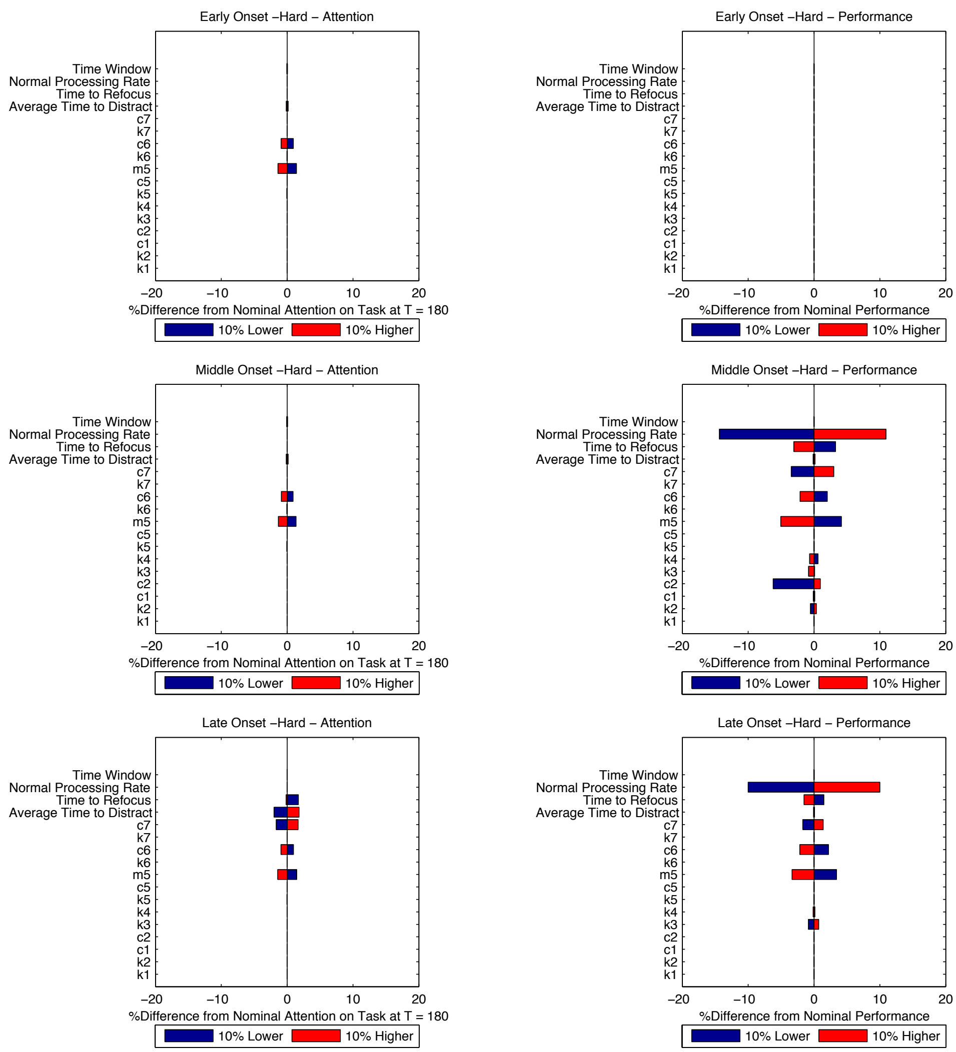


Figure E-10: Sensitivity Analysis – Hard Task

Other parameters that have relatively large impacts on model outputs are m_5, which corresponds to active fatigue; c_6, which corresponds to passive fatigue, and c_7, which corresponds to the effect of stress on attention. As discussed previously, active fatigue and passive fatigue influence the depletion of ECR, which then affects the attention distribution. Effect of stress on attention directly affects attention when task demand imposes a high level of stress. These variables are most closely related to attention in the PAL model in the causal structure. Hence, the output on attention is sensitive to their changes. These parameters are sensitive for both easy and hard tasks.

Summary

The sensitivity analyses show that the sensitivity of the model differs in different task settings. In the OPS-USERS task and the tracking task, the model is robust to changes in parameters in general. For the nuclear power plant monitoring task, the model is quite sensitive to changes in parameters, especially under the distraction condition. The common variables across the three task scenarios are identified and summarized in Table E-2. Normal Processing Rate is a sensitive variable for all three tasks. This makes sense because it directly relates to task processing in the casual loop structure in the PAL model. However, the high sensitivity also calls for improvement to the model in the stress and Yerkes-Dodson loops for uncertainty reduction. From parameter estimation aspect, we should devote more effort to the measure and estimation of this variable. The importance of Normal Processing Rate also indicates that it is a potential variable for performance improvement.

Table E-2: Common Variables Highlighted in the Sensitivity Analyses

Task Scenario	OPS-USERS	Nuclear Power Plant	Tracking Task
Active Fatigue	m_5		m_5
Passive Fatigue	c_6	c_6	c_6
Effect of Stress on Attention	c_7	c_7, k_7	c_7
Normal Processing Rate (NPR)	NPR	NPR	NPR

For attention, parameters relate to active fatigue, passive fatigue and effect of stress on attention are identified as sensitive variables. Active and passive fatigue connects with the depletion of ECR, and effect of stress on attention reflects the impact of task demand on attention distribution. They are the variables most closely related to attention distribution in the casual loop structure in the PAL model, and their changes have a large impact on model output. As a result, their values need to be estimated carefully. However, unlike Normal Processing Rate, these effects are much harder to assess and measure. The improvement on the modeling and estimation of these effects will depend upon further theoretical and empirical studies.

Appendix F. Impact of Sleep and Boredom Proneness Score

In the PAL model, the impact of sleep and boredom proneness was modeled together as a linear function in Chapter 6. In order to separate the impact of these two variables on performance, sleep in the previous two nights was varied from 10 to 21 hours with an increment of one hour, and boredom proneness score was varied from 1 to 17 with an increment of one. These ranges were selected based on the maximum and minimum values of experiment participants' characteristics in Chapter 6. Performance with different combinations of sleep and boredom proneness score was then predicted.

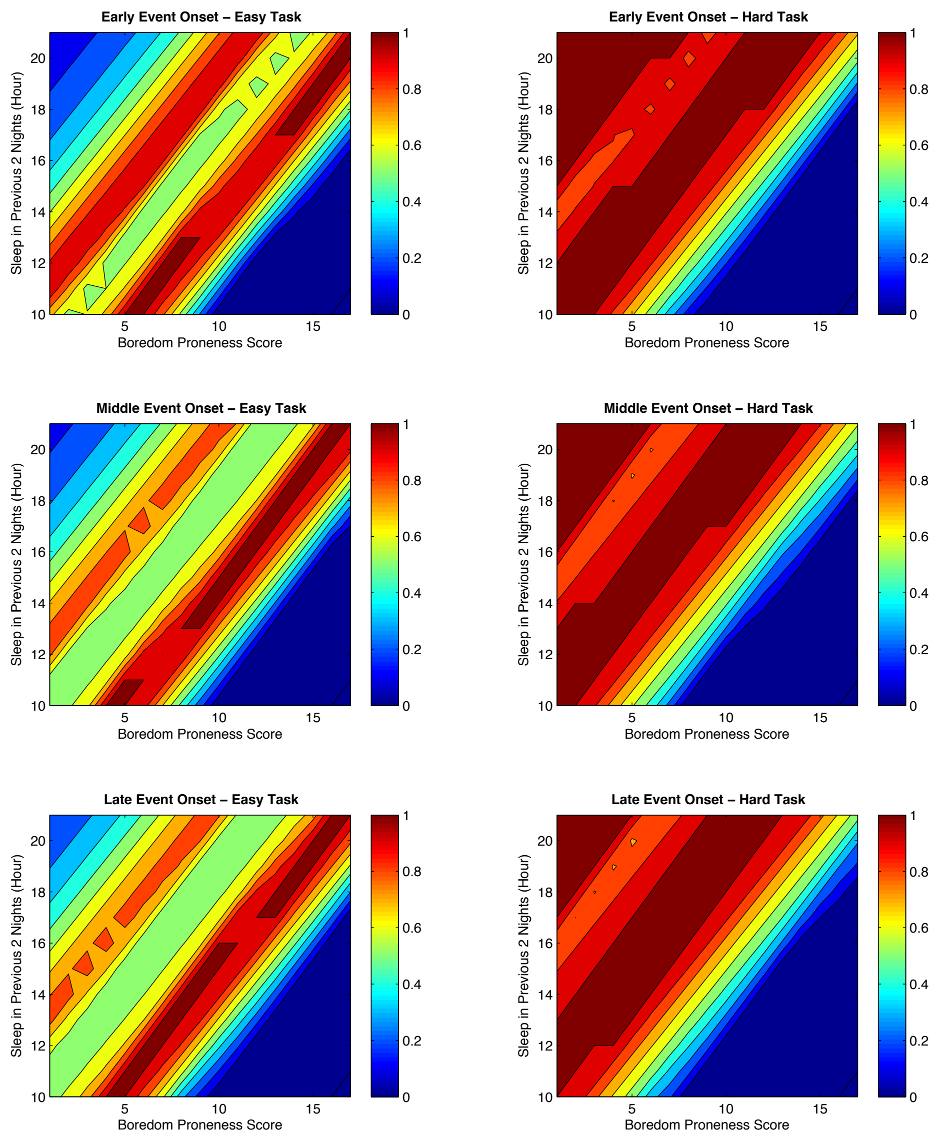


Figure F-1: Impact on of Sleep and BPS on Performance

In the easy task, there were three objects to be tracked. In the hard task, there were six objects to be tracked. To be consistent, performance was normalized as the percentage of objects tracked, thus the overall scale is 1. The results are presented in Figure F-1. The x axis denotes the boredom proneness score, and the y axis is hours of sleep in previous two nights. The color in each plot represents the performance ranging from zero to one. The expectation is that high level of sleep and low level of boredom proneness will result in better performance. In other words, the figures should show red colors in upper left corner and blue colors in lower right corner.

The impacts of sleep and boredom proneness were modeled as a linear function in the PAL model. With the hard task, high levels of sleep and low boredom proneness score result in better performance, which was expected. Capturing the performance changes with hard tasks is more critical in such applications, and the model is sufficiently accurate for this purpose. With the easy task, the model predicts a worse performance when the level of sleep is very high and the boredom proneness score is very low, as shown in the upper left corner in the plots. This contradicts with our expectation and reveals a limitation of the model. The reason for this counterintuitive result is that the attention level is high with such personal precursor values, resulting in a low stress level. The low stress level then causes a slower event processing rate, and ultimately less objects tracked. There are a few ways to solve this problem in the model. First, the representation of stress level needs to be changed. It is likely that actual stress level is higher than the current values used in the model when easy tasks are administered. The stress level is modeled as a ratio between Required Task Processing Rate and Individual Task Processing Rate. With a high level of individual capability and easy task, the stress level is low. However, it is possible that people feel aroused even with easy tasks. Second, the function that represents the effects of stress needs to be improved. Simple S-shaped curves currently used in the model as shown in Figure 3-8 may not be adequate. We leave these areas for future research.

Performance of all 30 participants in the experiment was plotted in Figure F-2 along with their sleep levels and boredom proneness scores to compare the simulation prediction with the experimental data. The x axis denotes the boredom proneness score, and the y axis is hours of sleep in the previous two nights. The color and size of the circles represents the performance ranging from 0% to 100%. Smaller circles and blue colors represent lower percentage of objects tracked. These performance data are from 30 participants under six experimental conditions with different event onset times and task difficulty. Event onset times and task difficulty affected the performance, but was not represented here. In this figure, the worst performance was achieved by a participant with high boredom proneness score (17 out of 24) and relatively low level of sleep (13 hours). Another observation is that the four participants that had over 18 hours of sleep all performed relatively well.

While sleep seems to have some impact on performance, the influence of boredom proneness is more ambiguous. The participant with low level of boredom proneness (1 out of 24) and low level of sleep (14 hours) had a bad performance. This suggests that sleep might be the main factor that affects task performance, and boredom proneness is not influential. In another study investigating the relationship between boredom proneness and sustained attention, it was found that boredom proneness was correlated with attention lapses ($r = 0.53$, $p = 0.01$), and attention-related cognitive errors ($r = 0.49$, $p = 0.01$) in everyday life as measured by two self-report questionnaires. However, boredom proneness is not correlated with sustained attention to response task (SART), for which participants need to respond to single digits presented on a computer screen, but to withhold a response to one particular digit ($r = 0.16$, $p = 0.28$) (Malkovsky et al. 2012). This suggests that boredom proneness, as a subjective measure, may not be a good predictor of task performance and should possibly be removed from the model.

With the limited data collected in our experiment, we could not derive a more accurate model to represent the individual differences adequately. This requires further research and a larger sample size. For attention, there are several task-based attention control tests that could provide proxy measures of individual differences including the Paced Auditory Serial-Attention Task (PASAT), Stroop and Test of Everyday Attention (TEA). For sustained attention, there are choice reaction time, continuous performance task, and go/no-go task (Shipstead et al. 2012). These tests may provide better assessment of individuals' attention abilities compared to boredom proneness score.

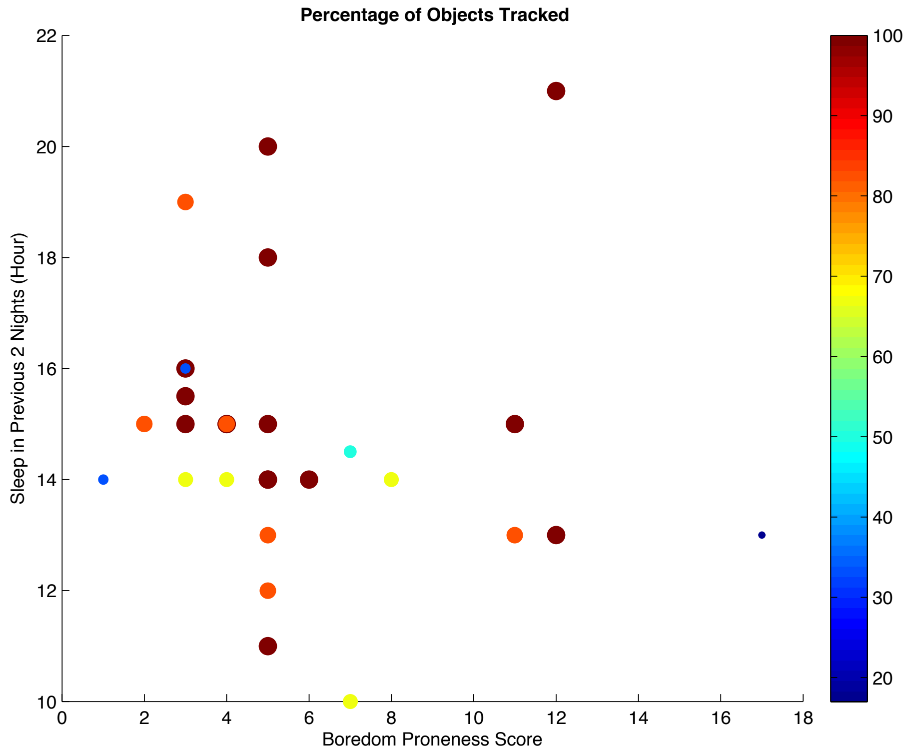


Figure F-2: Sleep, BPS and Performance in Experiment Data

Appendix G. Impact of Expertise

To examine the impact of expertise on task performance, two Monte Carlo simulations were run based on the nuclear power plant monitoring task described in Chapter 5. In the original model, Baseline Normal Processing Time was set to 20 minutes. Since the participants in the experiment had no previous experience with nuclear power plants, they were considered novices. To represent the processing time of novice in the Monte Carlo simulation, a normal distribution $X \sim N\ (\mu = 20, \sigma = 10)$ was used. In contrast, experts were represented by a normal distribution $X \sim N\ (\mu = 15, \sigma = 5)$, which means they could complete the tasks faster and were more consistent in their performance with smaller variance. These two distributions are visualized in Figure G-1. Random numbers were drawn from these two distributions with a upper bound of 40 minutes and a lower bound of 1 minute.

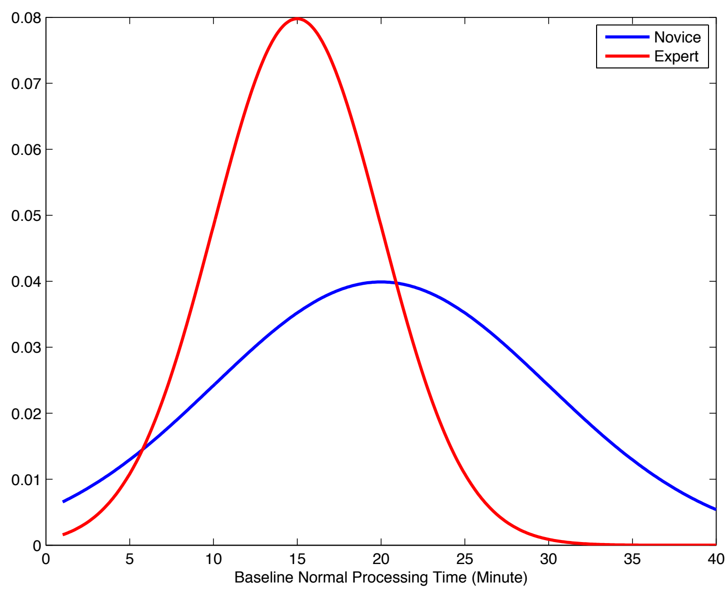


Figure G-1: Distributions for Processing Time

In Chapter 5, the experiment had six conditions: Operating Environment (Sterile, Distraction) x Event Onset Time (Early, Middle, Late). 200 simulations were run for each condition. The outputs on performance are presented in Figure G-2 as boxplots. Comparing novices with experts, it can be seen that experts have better (higher median value) and more consistent (tighter percentiles) performance. Comparing sterile with distraction condition, it can be seen that the percentiles of the performance are tighter in the sterile condition for both novices and experts. This means performance is less affected by variation in processing time when external distraction sources are restricted. In summary, the model's output on performance is consistent with our expectations for the differences between experts and novices, and thus while the experiment in Chapter 5 used novices, the model can handle expert operators as well.

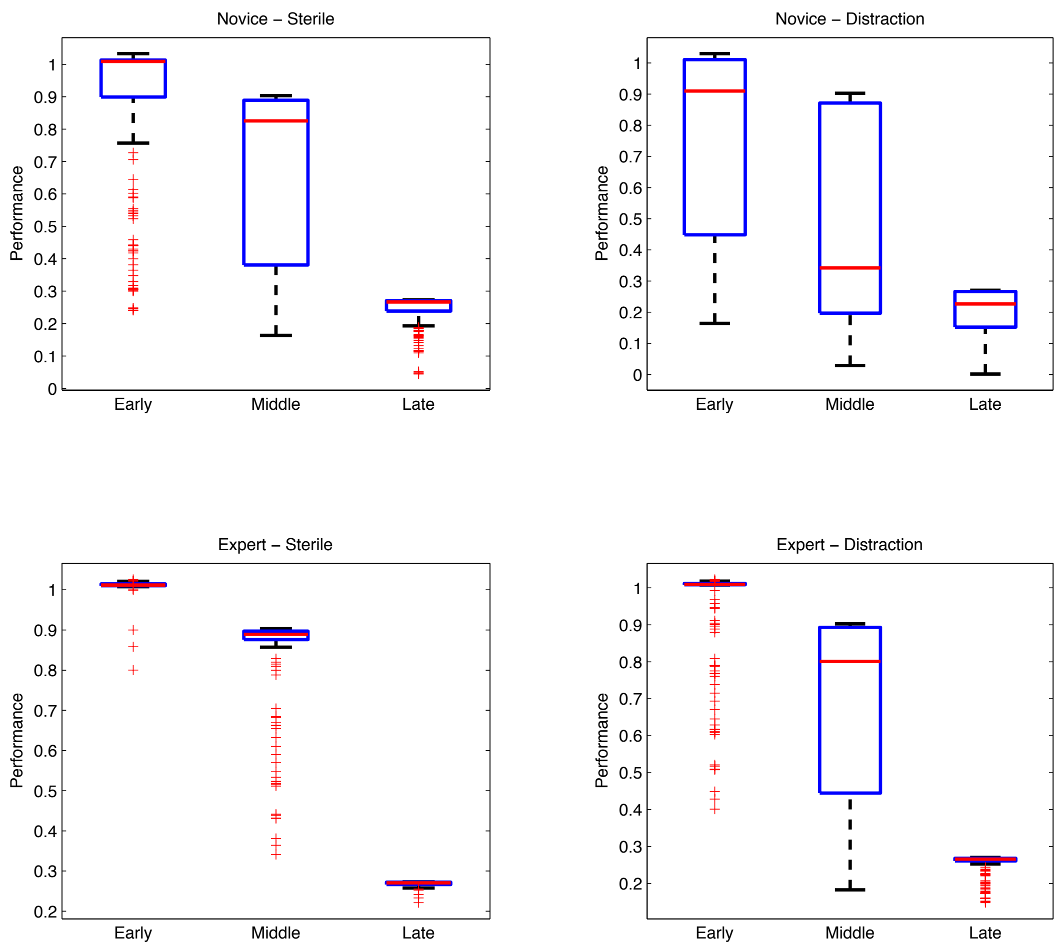


Figure G-2: Performance of Novices and Experts under Different Conditions

Reference

Aaronson, L. S., Teel, C. S., Cassmeyer, V., Neuberger, G. B., Pallikkathayil, L., Pierce, J., Press, A. N., Williams, P. D., and Wingate, A. (1999), Defining and measuring fatigue. *Image: the Journal of Nursing Scholarship, 31*(1), 45-50. doi: 10.1111/j.1547-5069.1999.tb00420.x

Abramson, E. E., and Stinson, S. G. (1977), Boredom and eating in obese and non-obese individuals. *Addictive Behaviors, 2*(4), 181-185. doi: http://dx.doi.org/10.1016/0306-4603(77)90015-6

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., Chang, R., and Jacob, R. J. K. (2014). *Dynamic difficulty using brain metrics of workload*. Paper presented at the Proceedings of the 32nd annual ACM conference on Human factors in computing systems, Toronto, Ontario, Canada.

Ahmed, S. (1990), Psychometric properties of the boredom proneness scale. *Perceptual and Motor Skills, 71*(3), 963-966.

Alves, E. E., and Kelsey, C. M. (2010), Combating vigilance decrement in a single-operator radar platform. *Ergonomics in Design: The Quarterly of Human Factors Applications, 18*(2), 6-9. doi: 10.1518/106480410x12737888532688

Ashton, M. C., and Lee, K. (2007), Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166. doi: 10.1177/1088868306294907

Atchley, P., and Chan, M. (2011), Potential benefits and costs of concurrent task engagement to maintain vigilance: A driving simulator investigation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(1), 3-12. doi: 10.1177/0018720810391215

Azarnoosh, M., Nasrabadi, A. M., Mohammadi, M. R., and Firoozabadi, M. (2012), Evaluating nonlinear variability of mental fatigue behavioral indices during long-term attentive task. *Complexity, 17*(6), 7-16. doi: 10.1002/cplx.21391

Azizi, N., Zolfaghari, S., and Liang, M. (2010), Modeling job rotation in manufacturing systems: The study of employee's boredom and skill variations. *International Journal of Production Economics, 123*(1), 69-85. doi: http://dx.doi.org/10.1016/j.ijpe.2009.07.010

Bainbridge, L. (1983), Ironies of automation. *Automatica, 19*(6), 775-779. doi: http://dx.doi.org/10.1016/0005-1098(83)90046-8

Ballard, J. C. (1996), Computerized assessment of sustained attention: A review of factors affecting vigilance performance. *Journal of Clinical and Experimental Neuropsychology, 18*(6), 843-863. doi: 10.1080/01688639608408307

Barata, G., Gama, S., Jorge, J., and Goncalves, D. (2013, September), *Engaging engineering students with gamification*. Paper presented at the Games and Virtual Worlds for Serious Applications (VS-GAMES), 2013 5th International Conference on.

Barmack, J. E. (1937), *Boredom and other factors in the physiology of mental effort: An exploratory study*. New York: Columbia university.

Barmack, J. E. (1939), Studies on the psychophysiology of boredom: Part I. The effect of 15 mgs. of benzedrine sulfate and 60 mgs. of ephedrine hydrochloride on blood pressure, report of boredom and other factors. *Journal of Experimental Psychology, 25*(5), 494-505. doi: 10.1037/h0054402

Baumeister, R. F. (2002), Ego depletion and self-control failure: An energy model of the self's executive function. *Self and Identity, 1*(2), 129-136. doi: 10.1080/152988602317319302

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998), Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252-1265. doi: 10.1037/0022-3514.74.5.1252

Becker, A. B., Warm, J. S., Dember, W. N., and Hancock, P. A. (1991), Effects of feedback on perceived workload in vigilance performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 35*(20), 1491-1494. doi: 10.1177/154193129103502015

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. (2007), EEG correlates of task engagement and mental workload in vigilance, learning, and memory Tasks. *Aviation, Space, and Environmental Medicine, 78*(5), B231-B244.

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2005), Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research, 25*(1), 107-116. doi: http://dx.doi.org/10.1016/j.cogbrainres.2005.04.011

Boyer, M. W. (2014), *The impact of rapid transitions in mental workload in a supervisory control setting.* Massachusetts Institute of Technology.

Brailsford, S. C., Desai, S. M., and Viana, J. (2010), *Towards the holy grail: combining system dynamics and discrete-event simulation in healthcare.* Paper presented at the Simulation Conference (WSC), Proceedings of the 2010 Winter.

Britton, A., and Shipley, M. J. (2010), Bored to death? *International Journal of Epidemiology, 39*(2), 370-371. doi: 10.1093/ije/dyp404

Bruursema, K., Kessler, S. R., and Spector, P. E. (2011), Bored employees misbehaving: The relationship between boredom and counterproductive work behaviour. *Work & Stress, 25*(2), 93-107. doi: 10.1080/02678373.2011.596670

Caggiano, D., and Parasuraman, R. (2004), The role of memory representation in the vigilance decrement. *Psychonomic Bulletin & Review, 11*(5), 932-937. doi: 10.3758/BF03196724

Cain, B. (2007). A review of the mental workload literature: DTIC Document.

Campagne, A., Pebayle, T., and Muzet, A. (2004), Correlation between driving errors and vigilance level: influence of the driver's age. *Physiology and Behavior, 80*(4), 515-524. doi: http://dx.doi.org/10.1016/j.physbeh.2003.10.004

Carpenito-Moyet, L. J. (2006), *Nursing diagnosis: Application to clinical practice.* Philadelphia: Lippincott Williams & Wilkins.

Casner, S. M., and Schooler, J. W. (2014), Thoughts in flight: Automation use and pilots' task-related and task-unrelated thought. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*(3), 433-442. doi: 10.1177/0018720813501550

Casner, S. M., and Schooler, J. W. (2015), Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and Cognition, 35*, 33-41. doi: http://dx.doi.org/10.1016/j.concog.2015.04.019

Chappelle, W., McDonald, K., Prince, L., Goodman, T., Ray-Sannerud, B. N., and Thompson, W. (2014), Assessment of occupational burnout in United States Air Force Predator/Reaper "drone" operators. *Military Psychology, 26*(5-6), 376-385. doi: 10.1037/mil0000046

Christoff, K., Ream, J. M., and Gabrieli, J. D. E. (2004), Neural basis of spontaneous thought processes. *Cortex, 40*(4–5), 623-630. doi: http://dx.doi.org/10.1016/S0010-9452(08)70158-8

Clare, A. S. (2013), *Modeling real-time human-automation collaborative scheduling of unmanned vehicles.* (Doctoral), Massachusetts Institute of Technology, Cambridge.

Costa, P. T., and McCrae, R. R. (1992), Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4*(1), 5-13. doi: http://dx.doi.org/10.1037/1040-3590.4.1.5
Coull, J. T., Frackowiak, R. S. J., and Frith, C. D. (1998), Monitoring for target objects: activation of right frontal and parietal cortices with increasing time on task. *Neuropsychologia, 36*(12), 1325-1334. doi: http://dx.doi.org/10.1016/S0028-3932(98)00035-9
Csikszentmihalyi, M. (1978). Attention and the holistic approach to behavior. In K. Pope & J. Singer (Eds.), *The Stream of Consciousness* (pp. 335-358). New York, NY: Plenum.
Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience *Flow and the foundations of positive psychology* (pp. 209-226). Netherlands: Springer Netherlands.
Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2014). Flow *Flow and the Foundations of Positive Psychology* (pp. 227-238): Springer Netherlands.
Csikszentmihalyi, M., and Larson, R. (1987), Validity and reliability of the experience-sampling method. *The Journal of nervous and mental disease, 175*(9), 526-536.
Culp, N. A. (2006), The relations of two facets of boredom proneness with the major dimensions of personality. *Personality and Individual Differences, 41*(6), 999-1007. doi: http://dx.doi.org/10.1016/j.paid.2006.02.018
Cummings, C., and Ryan, J. (2013), Shared Authority Concerns in Automated Driving Applications. *J Ergonomics S, 3*, 2.
Cummings, M. (2008), Of shadows and white scarves. *C4ISR Journal, 7*, 17-18.
Cummings, M., Gao, F., and Thornburg, K. M. (2015), Boredom in the Workplace: A New Look at an Old Problem. *Human Factors, In Press.*
Cummings, M. L., Clare, A., and Hart, C. (2010), The role of human-automation consensus in multiple unmanned vehicle scheduling. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52*(1), 17-27. doi: 10.1177/0018720810368674
Cummings, M. L., How, J. P., Whitten, A., and Toupet, O. (2012), The impact of human-automation collaboration in decentralized multiple unmanned vehicle control. *Proceedings of the IEEE, 100*(3), 660-671. doi: 10.1109/JPROC.2011.2174104
Cummings, M. L., Mastracchio, C., Thornburg, K. M., and Mkrtchyan, A. (2013), Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers, 25*(1), 34-47. doi: 10.1093/iwc/iws011
Cunningham, S., Scerbo, M. W., and Freeman, F. G. (2000), The electrocortical correlates of daydreaming during vigilance tasks. *Journal of Mental Imagery, 24*(1-2), 61-72.
Dahlen, E. R., Martin, R. C., Ragan, K., and Kuhlman, M. M. (2005), Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accident Analysis & Prevention, 37*(2), 341-348. doi: http://dx.doi.org/10.1016/j.aap.2004.10.006
Daniels, L. M., Tze, V. M. C., and Goetz, T. (2015), Examining boredom: Different causes for different coping profiles. *Learning and Individual Differences, 37*(0), 255-261. doi: http://dx.doi.org/10.1016/j.lindif.2014.11.004
Davies, D. R., and Parasuraman, R. (1982), *The psychology of vigilance*. London: Academic Press.
Denisco, R., Drummond, J., and Gravenstein, J. (1987), The effect of fatigue on the performance of a simulated anesthetic monitoring task. *Journal of Clinical Monitoring, 3*(1), 22-24. doi: 10.1007/BF00770879
Desmond, P. A., and Hancock, P. A. (2001). Active and passive fatigue states. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue* (pp. 455-465). Mahwah, NJ: Lawrence Erlbaum Associates.

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., and Zoladz, P. R. (2007), The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-Induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity, 2007*. doi: 10.1155/2007/60803

Dittmar, M. L., Warm, J. S., Dember, W. N., and Ricks, D. F. (1993), Sex differences in vigilance performance andperceived workload. *The Journal of General Psychology, 120*(3), 309-322. doi: 10.1080/00221309.1993.9711150

Drory, A. (1982), Individual differences in boredom proneness and task effectiveness at work. *Personnel Psychology, 35*(1), 141-151.

Dunn, N., and Williamson, A. (2011, July), *Monotony in the rail industry: The role of task demand in mitigating monotony-related effects on performance.* Paper presented at the Ergonomics Australia–HFESA 2011 Conference Edition.

Eastwood, J. D., Frischen, A., Fenske, M. J., and Smilek, D. (2012), The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science, 7*(5), 482-495. doi: 10.1177/1745691612456044

Eker, S., Slinger, J., van Daalen, E., and Yücel, G. (2014), Sensitivity analysis of graphical functions. *System Dynamics Review, 30*(3), 186-205. doi: 10.1002/sdr.1518

Eker, S., Slinger, J., and Yucel, G. (2011), *Investigating an automated method for the sensitivity analysis of functions.* Paper presented at the Proceedings of the 29th International Conference of the System Dynamics Society, Washington, USA, 25-29 July 2011.

Estes, S. (2015), The workload curve: Subjective mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(7), 1174-1187. doi: 10.1177/0018720815592752

Eysenck, M. W., and Calvo, M. G. (1992), Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*(6), 409-434. doi: 10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., and Calvo, M. G. (2007), Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336-353. doi: 10.1037/1528-3542.7.2.336

Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., and Eastwood, J. D. (2013), Development and validation of the multidimensional state boredom scale. *Assessment, 20*(1), 68-85. doi: 10.1177/1073191111421303

Farmer, R., and Sundberg, N. D. (1986), Boredom proneness--the development and correlates of a new scale. *Journal of Personality Assessment, 50*(1), 4-17.

Fei, G., Cummings, M. L., and Solovey, E. T. (2014), Modeling Teamwork in Supervisory Control of Multiple Robots. *Human-Machine Systems, IEEE Transactions on, 44*(4), 441-453. doi: 10.1109/THMS.2014.2312391

Fenichel, O. (1951). On the psychology of boredom. In D. Rapaport (Ed.), *Organization and pathology of thought: Selected sources* (pp. 349-361). New York, NY: Columbia University Press.

Finomore, V. S., Shaw, T. H., Warm, J. S., Matthews, G., and Boles, D. B. (2013), Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(6), 1044-1063. doi: 10.1177/0018720813484498

Fisher, C. D. (1993), Boredom at work: A neglected concept. *Human Relations, 46*(3), 395-417. doi: 10.1177/001872679304600305

Fisher, C. D. (1998), Effects of external and internal interruptions on boredom at work: two studies. *Journal of Organizational Behavior, 19*(5), 503-522. doi: 10.1002/(SICI)1099-1379(199809)19:5<503::AID-JOB854>3.0.CO;2-9
Fisher, S. (1986), *Stress and strategy*: Taylor & Francis.
Franco, D. (2007), *Fifty years of table functions.* Paper presented at the Proceedings of the 25th International Conference of the System Dynamics Society and 50th Anniversary Celebration.
Frederick-Recascino, C. M., and Hilscher, M. (2001, October), *Monitoring automated displays: Effects of and solutions for boredom.* Paper presented at the Digital Avionics Systems, 2001. DASC. 20th Conference.
Fuller, R. (2005), Towards a general theory of driver behaviour. *Accident Analysis & Prevention, 37*(3), 461-472. doi: http://dx.doi.org/10.1016/j.aap.2004.11.003
Gannes, L. (2014). Here's what it's like to go for a ride in google's robot car. from http://recode.net/2014/05/13/googles-self-driving-car-a-smooth-test-ride-but-a-long-road-ahead/
Gao, F., Cummings, M. L., and Solovey, E. T. (2014), Modeling teamwork in supervisory control of multiple robots. *Human-Machine Systems, IEEE Transactions on, 44*(4), 441-453. doi: 10.1109/THMS.2014.2312391
Geiselman, E. E., Johnson, C. M., and Buck, D. R. (2013), Flight deck automation: Invaluable collaborator or insidious enabler? *Ergonomics in Design: The Quarterly of Human Factors Applications, 21*(3), 22-26. doi: 10.1177/1064804613491268
Gershon, P., Ronen, A., Oron-Gilad, T., and Shinar, D. (2009), The effects of an interactive cognitive task (ICT) in suppressing fatigue symptoms in driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(1), 21-28. doi: http://dx.doi.org/10.1016/j.trf.2008.06.004
Giakoumis, D., Tzovaras, D., Moustakas, K., and Hassapis, G. (2011), Automatic recognition of boredom in video games using novel biosignal moment-based features. *Affective Computing, IEEE Transactions on, 2*(3), 119-133. doi: 10.1109/T-AFFC.2011.4
Gould, J. D., and Schaffer, A. M. Y. (1967), Eye-movement parameters in pattern recognition. *Journal of Experimental Psychology, 74*(2, Pt.1), 225-229. doi: http://dx.doi.org/10.1037/h0024559
Grech, M. R., Neal, A., Yeo, G., Humphreys, M., and Smith, S. (2009), An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology, 14*(3), 231-242. doi: http://dx.doi.org/10.1037/a0014952
Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., and Parasuraman, R. (2003), The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(3), 349-359. doi: 10.1518/hfes.45.3.349.27253
Gross, R. G., and Grossman, M. (2010), Executive resources. *Continuum (Minneapolis, Minn.), 16*(400), 140-152. doi: 10.1212/01.CON.0000368266.46038.0e
Haga, S. (1984), An experimental study of signal vigilance errors in train driving. *Ergonomics, 27*(7), 755-765. doi: 10.1080/00140138408963549
Hamilton, J. A., Haier, R. J., and Buchsbaum, M. S. (1984), Intrinsic enjoyment and boredom coping scales: Validation with personality, evoked potential and attention measures. *Personality and Individual Differences, 5*(2), 183-193. doi: http://dx.doi.org/10.1016/0191-8869(84)90050-3

Hamilton-Baillie, B. (2008), Shared space: Reconciling people, places and traffic. *Built Environment, 34*(2), 161-181. doi: 10.2148/benv.34.2.161

Hancock, P. A., and Desmond, P. A. (2001), *Stress, workload, and fatigue*: Lawrence Erlbaum Associates Publishers.

Hancock, P. A., and Warm, J. S. (2003), A dynamic model of stress and sustained attention. *Journal of Human Performance in Extreme Environments, 7*(1), 4.

Harris, M. B. (2000), Correlates and characteristics of boredom proneness and boredom. *Journal of Applied Social Psychology, 30*(3), 576-598. doi: 10.1111/j.1559-1816.2000.tb02497.x

Hart, C. S. (2010), *Assessing the impact of low workload in supervisory control of networked unmanned vehicles.* Massachusetts Institute of Technology, Cambridge.

Hart, S. G. (2006), Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. doi: 10.1177/154193120605000909

Harvey, J., Heslop, S., and Thorpe, N. (2010), The categorisation of drivers in relation to boredom. *Transportation Planning and Technology, 34*(1), 51-69. doi: 10.1080/03081060.2011.530829

Helton, W. S., and Warm, J. S. (2008), Signal salience and the mindlessness theory of vigilance. *Acta Psychologica, 129*(1), 18-25. doi: http://dx.doi.org/10.1016/j.actpsy.2008.04.002

Heslop, S., Harvey, J., Thorpe, N., and Mulley, C. (2009), Factors that comprise driver boredom and their relationships to preferred driving speed and demographic variables. *Transportation Planning and Technology, 33*(1), 75-89. doi: 10.1080/03081060903429389

Hill, A. B., and Perkins, R. E. (1985), Toward a model of boredom. *British Journal of Psychology, 76*(2), 235-240.

Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., and See, J. E. (1999), Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*(3), 365-372. doi: 10.1518/001872099779610987

Hockey, G. R. J. (1997), Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*(1–3), 73-93. doi: http://dx.doi.org/10.1016/S0301-0511(96)05223-4

Homer, J. B. (1985), Worker burnout: A dynamic model with implications for prevention and control. *System Dynamics Review, 1*(1), 42-62. doi: 10.1002/sdr.4260010105

Inzlicht, M., and Schmeichel, B. J. (2012), What Is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7*(5), 450-463. doi: 10.1177/1745691612454134

Inzlicht, M., Schmeichel, B. J., and Macrae, C. N. (2014), Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18*(3), 127-133. doi: http://dx.doi.org/10.1016/j.tics.2013.12.009

John, M., Smallman, H. S., and Manes, D. I. (2005), Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(3), 473-477. doi: 10.1177/154193120504900355

Johnson, A. W., Oman, C. M., Sheridan, T. B., and Duda, K. R. (2014, March), *Dynamic task allocation in operational systems: Issues, gaps, and recommendations.* Paper presented at the Aerospace Conference, 2014.

Kaber, D. B., and Endsley, M. R. (2004), The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153. doi: 10.1080/1463922021000054335

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

Kahneman, D. (2011), *Thinking, fast and slow*: Macmillan.

Kapoor, K., Subbian, K., Srivastava, J., and Schrater, P. (2015). *Just in time recommendations: Modeling the dynamics of boredom in activity streams*. Paper presented at the Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China.

Kara, D. (2013). Rio tinto rolls out ambitious, autonomous, mine of the future. from http://www.arisplex.com/analysis/rio-tinto-rolls-ambitious-autonomous-mine-future/

Kass, S., Vodanovich, S., and Callender, A. (2001), State-trait boredom: Relationship to absenteeism, tenure, and job satisfaction. *Journal of Business and Psychology, 16*(2), 317-327. doi: 10.1023/A:1011121503118

Kass, S. J., and Vodanovich, S. J. (1990), Boredom proneness: Its relationship to Type A behavior pattern and sensation seeking. *Psychology: A Journal of Human Behavior, 27*(3), 7-16.

Killingsworth, M. A., and Gilbert, D. T. (2010), A wandering mind is an unhappy mind. *Science, 330*(6006), 932-932. doi: 10.1126/science.1192439

Krueger, G. P. (1989), Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress, 3*(2), 129-141. doi: 10.1080/02678378908256939

Lal, S. K. L., and Craig, A. (2001), A critical review of the psychophysiology of driver fatigue. *Biological Psychology, 55*(3), 173-194. doi: http://dx.doi.org/10.1016/S0301-0511(00)00085-5

Langner, R., and Eickhoff, S. B. (2013), Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139*(4), 870-900. doi: http://dx.doi.org/10.1037/a0030694

Law, A., and Kelton, D. (2000), *Simulation Modeling and Analysis*: McGraw-Hill Higher Education.

Lee, K. A., Hicks, G., and Nino-Murcia, G. (1991), Validity and reliability of a scale to assess fatigue. *Psychiatry Research, 36*(3), 291-298. doi: http://dx.doi.org/10.1016/0165-1781(91)90027-M

LePera, N. (2011), The relationships between boredom proneness, mindfulness, anxiety, depression, and substance use. *The New School Psychology Bulletin, 8*(2), 15-25.

Lim, J., Wu, W.-c., Wang, J., Detre, J. A., Dinges, D. F., and Rao, H. (2010), Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *Neuroimage, 49*(4), 3426-3435. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.11.020

Locke, E. A., and Bryan, J. F. (1967), Performance goals as determinants of level of performance and boredom. *Journal of Applied Psychology, 51*(2), 120-130. doi: 10.1037/h0024471

Logan, G. D. (1985), Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology/Revue canadienne de psychologie, 39*(2), 367-386. doi: http://dx.doi.org/10.1037/h0080066

Loh, S., Lamond, N., Dorrian, J., Roach, G., and Dawson, D. (2004), The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers, 36*(2), 339-346. doi: 10.3758/BF03195580

Lorist, M. M., Klein, M., Nieuwenhuis, S., de Jong, R., Mulder, G., and Meijman, T. F. (2000), Mental fatigue and task control: Planning and preparation. *Psychophysiology, 37*(5), 614-625. doi: 10.1111/1469-8986.3750614

Loukidou, L., Loan-Clarke, J., and Daniels, K. (2009), Boredom in the workplace: More than monotonous tasks. *International Journal of Management Reviews, 11*(4), 381-405. doi: 10.1111/j.1468-2370.2009.00267.x

Loukopoulos, L. D., Dismukes, R. K., and Barshi, I. (2012), *The multitasking myth: Handling complexity in real-world operations*. Farnham, UK: Ashgate Publishing, Ltd.

MacDonald, W. (2003), The impact of job demands and workload on stress and fatigue. *Australian Psychologist, 38*(2), 102-117.

Mackworth, N. H. (1957), Some factors affecting vigilance. *Advancement of Science, 53*, 389-393.

Malkovsky, E., Merrifield, C., Goldberg, Y., and Danckert, J. (2012), Exploring the relationship between boredom and sustained attention. *Experimental Brain Research, 221*(1), 59-67. doi: 10.1007/s00221-012-3147-z

Manly, T., Robertson, I. H., Galloway, M., and Hawkins, K. (1999), The absent mind: further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661-670. doi: http://dx.doi.org/10.1016/S0028-3932(98)00127-4

Martella, D., Casagrande, M., and Lupiáñez, J. (2011), Alerting, orienting and executive control: the effects of sleep deprivation on attentional networks. *Experimental Brain Research, 210*(1), 81-89. doi: 10.1007/s00221-011-2605-3

Matic, A., Pielot, M., and Oliver, N. (2015). *Boredom-computer interaction: boredom proneness and the use of smartphone*. Paper presented at the Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan.

Matthews, G., and Campbell, S. E. (1998), Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(11), 821-825. doi: 10.1177/154193129804201111

Matthews, G., Campbell, S. E., Desmond, P. A., Huggins, J., Falconer, S., and Joyner, A. (1999), Assessment of task-induced state change: Stress, fatigue and workload components. *Automation technology and human performance: Current research and trends*, 199-203.

Matthews, G., and Desmond, P. A. (2002), Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A, 55*(2), 659-686. doi: 10.1080/02724980143000505

Matthews, G., Hancock, P. A., Desmond, P. A., and Neubauer, C. (2012), *The handbook of operator fatigue*. Burlington, VT: Ashgate Publishing Company.

Matthews, G., Warm, J. S., Reinerman, L. E., Langheim, L. K., and Saxby, D. J. (2010), *Task engagement, attention, and executive control*: Springer.

Matthews, G., Warm, J. S., Reinerman-Jones, L. E., Langheim, L. K., Washburn, D. A., and Tripp, L. (2010), Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *Journal of Experimental Psychology: Applied, 16*(2), 187-203. doi: http://dx.doi.org/10.1037/a0019572

Matthews, G., Warm, J. S., Shaw, T. H., and Finomore, V. S. (2010), A multivariate test battery for predicting vigilance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(14), 1072-1076. doi: 10.1177/154193121005401405

Matthews, G., Warm, J. S., Shaw, T. H., and Finomore, V. S. (2014), Predicting battlefield vigilance: a multivariate approach to assessment of attentional resources. *Ergonomics, 57*(6), 856-875. doi: 10.1080/00140139.2014.899630

McBain, W. N. (1970), Arousal, monotony, and accidents in line driving. *Journal of Applied Psychology, 54*(6), 509-519. doi: http://dx.doi.org/10.1037/h0030144

McDowall, R. J. S., and Wells, H. M. (1927), The physiology of monotony. *British Medical Journal, 1*(3452), 414-415.

McIntosh, J., MacDonald, F., and McKeganey, N. (2005), Pre-teenage children's experiences of drug use. *International Journal of Drug Policy, 16*(1), 37-45. doi: http://dx.doi.org/10.1016/j.drugpo.2004.06.007

McVay, J. C., and Kane, M. J. (2010), Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136*(2), 188-197. doi: http://dx.doi.org/10.1037/a0018298

Merat, N., Jamson, A. H., Lai, F. C. H., and Carsten, O. (2012), Highly automated driving, secondary task performance, and driver state. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(5), 762-771. doi: 10.1177/0018720812442087

Mercer-Lynn, K. B., Flora, D. B., Fahlman, S. A., and Eastwood, J. D. (2013), The measurement of boredom: Differences between existing self-report scales. *Assessment, 20*(5), 585-596. doi: 10.1177/1073191111408229

Merrifield, C., and Danckert, J. (2014), Characterizing the psychophysiological signature of boredom. *Experimental Brain Research, 232*(2), 481-491. doi: 10.1007/s00221-013-3755-2

Mikulas, W. L., and Vodanovich, S. J. (1993), The essence of boredom. *The Psychological Record, 43*(1), 3-12.

Mkrtchyan, A. A. (2011), *Modeling operator performance in low task load supervisory domains.* Massachusetts Institute of Technology.

Mkrtchyan, A. A., Macbeth, J. C., Solovey, E. T., Ryan, J. C., and Cummings, M. L. (2012), Using variable-rate alerting to counter boredom in human supervisory control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 1441-1445. doi: 10.1177/1071181312561406

Monsell, S. (2003), Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140. doi: http://dx.doi.org/10.1016/S1364-6613(03)00028-7

Moray, N. (2013), *Mental workload: Its theory and measurement* (Vol. 8): Springer Science & Business Media.

Morcom, A. M., and Fletcher, P. C. (2007), Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage, 37*(4), 1073-1082. doi: http://dx.doi.org/10.1016/j.neuroimage.2006.09.013

Morecroft, J., and Robinson, S. (2005), *Explaining puzzling dynamics: comparing the use of system dynamics and discrete-event simulation.* Paper presented at the Proceedings of the 23rd International Conference of the System Dynamics Society.

Morecroft, J. D. W. (1985), Rationality in the analysis of behavioral simulation models. *Management Science, 31*(7), 900-916. doi: 10.2307/2631605

Movie-watching air traffic controller suspended. (2011). from http://www.cbsnews.com/stories/2011/04/18/national/main20055158.shtml

Mumaw, R. J., Roth, E. M., Vicente, K. J., and Burns, C. M. (2000), There is more to monitoring a nuclear power plant than meets the eye. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 42*(1), 36-55. doi: 10.1518/001872000779656651

Muraven, M., and Slessareva, E. (2003), Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29*(7), 894-906. doi: 10.1177/0146167203029007008

Muraven, M., Tice, D. M., and Baumeister, R. F. (1998), Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*(3), 774-789. doi: 10.1037/0022-3514.74.3.774

Nehme, C. E. (2009), *Modeling Human Supervisory Control in Heterogeneous Unmanned Vehicle Systems.* (PhD), Massachusetts Institute of Technology, Cambridge.

Nett, U. E., Goetz, T., and Hall, N. C. (2011), Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology, 36*(1), 49-59. doi: http://dx.doi.org/10.1016/j.cedpsych.2010.10.003

Nietzsche, F. (1878), *Human, all too human: A book for free spirits.*

Norman, D. A. (2004), *Emotional design: Why we love (or hate) everyday things*: Basic books.

Northwest airlines flight 188. (2009). from http://www.faa.gov/data_research/accident_incident/2009-10-23/

O'Hanlon, J. F. (1981), Boredom: Practical consequences and a theory. *Acta Psychologica, 49*(1), 53-82. doi: http://dx.doi.org/10.1016/0001-6918(81)90033-0

Onnasch, L., Wickens, C. D., Li, H., and Manzey, D. (2014), Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*(3), 476-488. doi: 10.1177/0018720813501549

Ophir, E., Nass, C., and Wagner, A. D. (2009), Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*(37), 15583-15587. doi: 10.1073/pnas.0903620106

Oron-Gilad, T., Ronen, A., and Shinar, D. (2008), Alertness maintaining tasks (AMTs) while driving. *Accident Analysis & Prevention, 40*(3), 851-860. doi: http://dx.doi.org/10.1016/j.aap.2007.09.026

Oron-Gilad, T., and Shinar, D. (2000), Driver fatigue among military truck drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 3*(4), 195-209. doi: http://dx.doi.org/10.1016/S1369-8478(01)00004-3

Owens, B. D., Leveson, N. G., and Hoffman, J. A. (2011), Procedure rework: a dynamic process with implications for the "rework cycle" and "disaster dynamics". *System Dynamics Review, 27*(3), 244-269. doi: 10.1002/sdr.464

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance* (pp. 1-39). Oxford, England: John Wiley & Sons.

Parasuraman, R., and Davies, D. R. (1977). A taxonomic analysis of vigilance performance. In R. R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates.* (Vol. 3, pp. 559-574). New York, NY: Plenum.

Parasuraman, R., and Manzey, D. H. (2010), Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52*(3), 381-410. doi: 10.1177/0018720810376055

Parasuraman, R., Mouloua, M., and Molloy, R. (1996), Effects of adaptive task allocation on monitoring of automated systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(4), 665-679. doi: 10.1518/001872096778827279

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000), A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 30*(3), 286-297. doi: 10.1109/3468.844354

Patten, C. J. D., Kircher, A., Östlund, J., and Nilsson, L. (2004), Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis & Prevention, 36*(3), 341-350. doi: http://dx.doi.org/10.1016/S0001-4575(03)00014-9

Pattyn, N., Neyt, X., Henderickx, D., and Soetens, E. (2008), Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior, 93*(1–2), 369-378. doi: http://dx.doi.org/10.1016/j.physbeh.2007.09.016

Petersen, S. E., and Posner, M. I. (2012), The attention system of the human brain: 20 years after. *Annual review of neuroscience, 35*, 73-89. doi: 10.1146/annurev-neuro-062111-150525

Pielot, M., Dingler, T., Pedro, J. S., and Oliver, N. (2015). *When attention is not scarce - detecting boredom from mobile phone usage*. Paper presented at the Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan.

Pop, V. L., Stearman, E. J., Kazi, S., and Durso, F. T. (2012), Using engagement to negate vigilance decrements in the NextGen environment. *International Journal of Human-Computer Interaction, 28*(2), 99-106. doi: 10.1080/10447318.2012.634759

Posner, M. I., and Petersen, S. E. (1990), The attention system of the human brain. *Annual review of neuroscience, 13*, 25-42. doi: 10.1146/annurev.ne.13.030190.000325

Prinzel, L. J., and Freeman, F. G. (1997), Task-specific sex differences in vigilance performance: Subjective workload and boredom. *Perceptual and Motor Skills, 85*(3f), 1195-1202. doi: 10.2466/pms.1997.85.3f.1195

Rahmandad, H., and Sterman, J. (2008), Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science, 54*(5), 998-1014. doi: doi:10.1287/mnsc.1070.0787

Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006), An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications, 9*(1), 58-69. doi: 10.1007/s10044-006-0025-y

Reinerman-Jones, L. E., Matthews, G., Langheim, L. K., and Warm, J. S. (2010), Selection for vigilance assignments: a review and proposed new direction. *Theoretical Issues in Ergonomics Science, 12*(4), 273-296. doi: 10.1080/14639221003622620

Rudolph, J. W., Morrison, J. B., and Carroll, J. S. (2009), The dynamics of action-oriented problem solving: Linking interpretation and choice. *Academy of Management Review, 34*(4), 733-756.

Rudolph, J. W., and Repenning, N. P. (2002), Disaster dynamics: Understanding the role of quantity in organizational collapse. *Administrative Science Quarterly, 47*(1), 1-30. doi: 10.2307/3094889

Russ, M., and Crews, D. E. (2014), A survey of multitasking behaviors in organizations. *International Journal of Human Resource Studies, 4*(1), Pages 137-153.

Russell, J. A. (1980), A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. doi: 10.1037/h0077714

Ryan, J. C. (2014), *Evaluating safety protocols for manned-unmanned environments through agent-based simulation.* (Doctoral), Massachusetts Institute of Technology.

Salvendy, G. (2012), *Handbook of human factors and ergonomics*: John Wiley & Sons.

Sawin, D. A., and Scerbo, M. W. (1994), Vigilance: How to do it and who should do it. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(19), 1312-1316. doi: 10.1177/154193129403801914

Sawin, D. A., and Scerbo, M. W. (1995), Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(4), 752-765. doi: 10.1518/001872095778995616

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., and Neubauer, C. (2013), Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied, 19*(4), 287-300. doi: http://dx.doi.org/10.1037/a0034386

Scerbo, M. W. (1998a), Sources of stress and boredom in vigilance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(10), 764-768. doi: 10.1177/154193129804201024

Scerbo, M. W. (1998b). What's so boring about vigilance? In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 145-166). Washington, DC: American Psychological Association.

Scerbo, M. W., Greenwald, C. Q., and Sawin, D. A. (1993), The effects of subject-controlled pacing and task type on sustained attention and subjective workload. *The Journal of General Psychology, 120*(3), 293-307. doi: 10.1080/00221309.1993.9711149

Schmeichel, B. J. (2007), Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General, 136*(2), 241-255. doi: http://dx.doi.org/10.1037/0096-3445.136.2.241

Schroeder, D. J., Touchstone, R. M., Stern, J. A., Stoliarov, N., and Thackray, R. I. (1994). Maintaining vigilance on a simulated ATC monitoring task across repeated sessions: DTIC Document.

Schroeter, R., Oxtoby, J., and Johnson, D. (2014, September), *AR and gamification concepts to reduce driver boredom and risk taking behaviours.* Paper presented at the Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA.

Scott, S. D., Mercier, S., Cummings, M. L., and Wang, E. (2006), Assisting interruption recovery in supervisory control of multiple UAVs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(5), 699-703. doi: 10.1177/154193120605000518

Shaw, T. H., Matthews, G., Warm, J. S., Finomore, V. S., Silverman, L., and Costa, P. T. (2010), Individual differences in vigilance: Personality, ability and states of stress. *Journal of Research in Personality, 44*(3), 297-308. doi: http://dx.doi.org/10.1016/j.jrp.2010.02.007

Sheridan, T. B., Vámos, T., and Aida, S. (1983), Adapting automation to man, culture and society. *Automatica, 19*(6), 605-612. doi: http://dx.doi.org/10.1016/0005-1098(83)90024-9

Shipstead, Z., Redick, T. S., and Engle, R. W. (2012), Is working memory training effective? *Psychological Bulletin, 138*(4), 628-654. doi: 10.1037/a0027473

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., and Obonsawin, M. (2004), Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13*(4), 657-690. doi: http://dx.doi.org/10.1016/j.concog.2004.06.003

Smallwood, J., Nind, L., and O'Connor, R. C. (2009), When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition, 18*(1), 118-125. doi: http://dx.doi.org/10.1016/j.concog.2008.11.004

Smallwood, J., and Schooler, J. W. (2006), The restless mind. *Psychological Bulletin, 132*(6), 946-958. doi: http://dx.doi.org/10.1037/0033-2909.132.6.946

Smith, R. P. (1981), Boredom: A review. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 23*(3), 329-340. doi: 10.1177/001872088102300308

Sommers, J., and Vodanovich, S. J. (2000), Boredom proneness: Its relationship to psychological-and physical-health symptoms. *Journal of clinical psychology, 56*(1), 149-155.

Stager, P., Hameluck, D., and Jubis, R. (1989, October), *Underlying factors in air traffic control incidents.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Denver, Colorado.

Sterman, J. D. (1984), Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica, 10*(2), 51-66.

Sterman, J. D. (2000), *Business dynamics: systems thinking and modeling for a complex world* (Vol. 19): Irwin/McGraw-Hill Boston.

Straussberger, S., and Schaefer, D. (2007), Monotony in air traffic control. *Air traffic control quarterly, 15*(3), 183-207.

Sundberg, N. D., Latkin, C. A., Farmer, R. F., and Saoud, J. (1991), Boredom in young adults: Gender and cultural comparisons. *Journal of Cross-Cultural Psychology, 22*(2), 209-223. doi: 10.1177/0022022191222003

Sweetser, A. (1999), *A comparison of system dynamics (SD) and discrete event simulation (DES).* Paper presented at the 17th International Conference of the System Dynamics Society.

Szalma, J. L., and Matthews, G. (2015). Motivation and emotion in sustained attention. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The Cambridge Handbook of Applied Perception Research* (Vol. 1, pp. 218-240). New York, NY: Cambridge University Press.

Szalma, J. L., Warm, J. S., Matthews, G., Dember, W. N., Weiler, E. M., Meier, A., and Eggemeier, F. T. (2004), Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors, 46*(2), 219-233.

Teaching 'the least boring job'. (2006). from http://news.bbc.co.uk/2/hi/uk_news/education/5217736.stm

Teigen, K. H. (1994), Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology, 4*(4), 525-547. doi: 10.1177/0959354394044004

Thackray, R. I. (1981), The stress of boredom and monotony: A consideration of the evidence. *Psychosomatic Medicine, 43*(2), 165-176.

Thackray, R. I., Jones, K. N., and Touchstone, R. M. (1974), Personality and physiological correlates of performance decrement on a monotonous task requiring sustained attention. *British Journal of Psychology, 65*(3), 351-358.

Thiffault, P., and Bergeron, J. (2003), Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention, 35*(3), 381-391. doi: http://dx.doi.org/10.1016/S0001-4575(02)00014-3

Thompson, W. T., Lopez, N., Hickey, P., DaLuz, C., Caldwell, J. L., and Tvaryanas, A. P. (2006). Effects of shift work and sustained operations: Operator performance in remotely piloted aircraft (OP-REPAIR): DTIC Document.

Thomson, D. R., Besner, D., and Smilek, D. (2015), A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science, 10*(1), 82-96. doi: 10.1177/1745691614556681

Thornburg, K. M., Peterse, H. P. M., and Liu, A. M. (2012), Operator performance in long duration control operations: Switching from low to high task load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 2002-2005. doi: 10.1177/1071181312561418

Thornburg, K. M., Peterse, H. P. M., Liu, A. M., and Oman, C. (2011). Operator performace in long duration, low task load control operations. http://www.web.mit.edu/aeroastro/labs/halab/papers/HAL2011-04.pdf

Tyler, D. M., and Halcomb, C. G. (1974), Monitoring Performance with a time-shared encoding task. *Perceptual and Motor Skills, 38*(2), 383-386. doi: 10.2466/pms.1974.38.2.383

Uys, P. (1984), An analytical form for table functions used in system dynamics. *Dynamica, 10*(2), 81-83.

van der Linden, D., Frese, M., and Meijman, T. F. (2003), Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica, 113*(1), 45-65. doi: http://dx.doi.org/10.1016/S0001-6918(02)00150-6

Van Dongen, H. P. A., and Belenky, G. (2009), Individual differences in vulnerability to sleep loss in the work environment. *Industrial Health, 47*(5), 518-526. doi: 10.2486/indhealth.47.518

Vodanovich, S. J. (2003), Psychometric measures of boredom: A review of the literature. *The Journal of Psychology, 137*(6), 569-595. doi: 10.1080/00223980309600636

Vodanovich, S. J., and Kass, S. J. (1990a), Age and gender differences in boredom proneness. *Journal of Social Behavior and Personality, 5*(4), 297-307.

Vodanovich, S. J., and Kass, S. J. (1990b), A factor analytic study of the boredom proneness scale. *Journal of Personality Assessment, 55*(1-2), 115-123. doi: 10.1080/00223891.1990.9674051

Vodanovich, S. J., Verner, K. M., and Gilbride, T. V. (1991), Boredom proneness: Its relationship to positive and negative affect. *Psychological Reports, 69*(3f), 1139-1146. doi: 10.2466/pr0.1991.69.3f.1139

Vodanovich, S. J., Wallace, J. C., and Kass, S. J. (2005), A confirmatory approach to the factor structure of the boredom proneness scale: Evidence for a two-factor short form. *Journal of Personality Assessment, 85*(3), 295-303. doi: 10.1207/s15327752jpa8503_05

Warm, J. S., Dember, W. N., and Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183-200). Hillsdale, NJ: Lawrence Erlbaum Associates.

Warm, J. S., Finomore, V. S., Vidulich, M. A., and Funke, M. E. (2015). Vigilance: A perceptual challenge. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The Cambridge Handbook of Applied Perception Research* (Vol. 1, pp. 241-283). New York, NY: Cambridge University Press.

Warm, J. S., Matthews, G., and Finomore, V. S., Jr. (2008). Vigilance, workload, and stress. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 115-141). Burlington, VT: Ashgate.

Warm, J. S., Parasuraman, R., and Matthews, G. (2008), Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 433-441. doi: 10.1518/001872008x312152

Washburn, D. A., Latzman, R. D., Schwartz, N. B., and Bramlett, J. (2015). Attention as a cause and an effect of perception. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The Cambridge Handbook of Applied Perception Research* (Vol. 1, pp. 126-147). New York, NY: Cambridge University Press.

Watt, J., and Hargis, M. (2010), Boredom proneness: Its relationship with subjective underemployment, perceived organizational support, and job performance. *Journal of Business and Psychology, 25*(1), 163-174. doi: 10.1007/s10869-009-9138-9

Weinger, M. (1999), Vigilance, boredom, and sleepiness. *Journal of Clinical Monitoring and Computing, 15*(7-8), 549-552. doi: 10.1023/A:1009993614060

Weinger, M., and Englund, C. (1990), Ergonomic and human factors affecting anesthetic vigilance and monitoring performance in the operating room environment. *Anesthesiology, 73*(5), 995-1021.

Wickens, C. D., Lee, J. D., Liu, Y., and Gordon-Becker, S. E. (2011), *An introduction to human factors engineering* (2 ed.). Upper Saddle River, NJ: Prentice Hall.

Wiener, E. L., and Nagel, D. C. (1988), *Human factors in aviation*. San Diego, CA: Academic Press.

Wierwille, W. W., Rahimi, M., and Casali, J. G. (1985), Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 27*(5), 489-502. doi: 10.1177/001872088502700501

Wiggins, M. W. (2011), Vigilance decrement during a simulated general aviation flight. *Applied Cognitive Psychology, 25*(2), 229-235. doi: 10.1002/acp.1668

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. L., and Shaked, A. (2014), Just think: The challenges of the disengaged mind. *Science, 345*(6192), 75-77. doi: 10.1126/science.1250830

Yerkes, R. M., and Dodson, J. D. (1908), The relation of strength of stimulus to rapidity of habit‐formation. *Journal of comparative neurology and psychology, 18*(5), 459-482.

Zhou, S. X. (2010), *Gratifications, loneliness, leisure boredom and self-esteem as predictors of SNS-game addiction and usage pattern among Chinese college students.* The Chinese University of Hong Kong.

Zuckerman, M. (1971), Dimensions of sensation seeking. *Journal of Consulting and Clinical Psychology, 36*(1), 45-52. doi: http://dx.doi.org/10.1037/h0030478