Assessing the Impact of Low Workload in Supervisory Control of Networked Unmanned Vehicles

by

Christin S. Hart

B.S. Astronautical Engineering
United States Air Force Academy, 2008

Submitted to the Department of Aeronautics and Astronautics in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author...
Department of Aeronautics and Astronautics
May 21, 2010

Certified by ....
Mary L. Cummings
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by ....
Eytan H. Modiano
Associate Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

Assessing the Impact of Low Workload in Supervisory Control of Networked Unmanned Vehicles

by

Christin S. Hart

Submitted to the Department of Aeronautics and Astronautics on May 21, 2010
in partial fulfillment of the requirements for the degree of
Master of Science in Aeronautics and Astronautics.

Abstract

This research investigated the effects of prolonged low workload on operator performance in the context of controlling a network of unmanned vehicles (UxVs) in a search, track, and destroy mission with the assistance of an automated planner. In addition, this research focused on assessing the physical, social, and cognitive coping mechanisms that operators rely upon during prolonged low workload missions. An experiment was conducted to collect data for researching the impact of low workload in human supervisory control of networked, heterogeneous UxVs. This research showed that performance was not necessarily affected at the low end of the workload spectrum, especially in the context of human supervisory control of networked UxVs. Given varying levels of low taskload, operators tended to gravitate toward a common total utilization (percent busy time) that was well above the required utilization. The boredom due to the low taskload environment caused operators to spend the majority of their time distracted; to a lesser degree, operators were more directed than divided in terms of attention. More directed attention predicted higher operator performance, especially in the tracking portion of the mission. Higher utilization predicted improved operator performance in search and destroy tasks, but hindered the automation's ability to track targets. Video gaming experience was a detriment to destroying hostile targets in this long duration, low workload mission involving human supervisory control of networked UxVs. Vigilance, shown by a decrement in amount of directed attention per hour, decreased over the course of the mission duration. Top performers had higher directed attention and coped with the boredom through extreme focus or use of switching times to stay engaged in the mission. In comparison to a moderate workload study, participants in this low workload experiment performed both better and worse. Low workload did not necessarily cause a drop in operator performance.

Acknowledgments

First, I owe my deepest gratitude to my advisor, Missy Cummings. Thank you for bringing me into your amazing lab and for guiding me through my research. I could not have asked for more one-on-one support. Thank you for loving your lab. You are a true leader to us. Most of all I want to thank you for being such a tremendous mentor in my life, especially as a military and professional role model. How I want to follow in your footsteps! Thanks for teaching me the importance of making smart career choices and for opening my mind to the vast opportunities the world has in store for me.

Thanks to Kris Thornburg, Emily Levesque, Americo Caves, and my family for providing feedback on this thesis.

I extend my gratitude to the Office of Naval Research for funding this research.

Thank you, Andrew, for being my closest friend and research partner. Everything I accomplished was with your help.

Thanks, Dan, for always helping me. I owe everything to your coding talent, from making my simulation to turning my raw data into new metrics.

Thanks to my undergraduates, Morris Vanegas and Vicki Crosson. I could not have gotten through the eternity-long experiments and videos without you.

Thank you to my colleagues at Aurora Flight Sciences and the Aerospace Controls Lab.

Thanks to my labmates: Geoff, Mariela, Anna, Birsen, Armen, Yves, Luca, Fish, Dave, Paul, Carine, Jackie, Farzan, Carl, Brian & Tuco, Jason, Ryan, JC, Ian, Pierre, Niek, Tony, Yale, Natasha, Sylvain, Jodyann, Suhail, Americo, Sally, Ping, Charlotte, Mark, Kris, Thomas, & Clotilde. You are my true friends and teammates.

To my mom and dad, you are the shoulders that I stand on. Your unending love fills me with joy and motivation. Thank you for being so proud of me. I am ever grateful for the countless hours of prayers you have offered up for me.

To my brothers, Ryan and Jonathan, thanks for being so loving and supportive to me our entire lives. You are both my inspiration and the model of Christ's love in my life.

Most importantly, I give thanks to God who makes all things possible. Philippians 4:13

Thank you, David, for all your love and for helping me to finish strong at MIT.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Expeditionary networks of unmanned vehicles (UxVs) are envisioned to be key resources in persistent surveillance [1]. These heterogeneous, unmanned vehicles will be highly autonomous. They will collaborate as a network of smart robots, equipped with onboard computers and communication devices. The mission environment will be dynamic and time-sensitive, requiring real-time, automated schedule replanning. A pictoral representation of the vision for networked UxVs is shown in Figure 1.

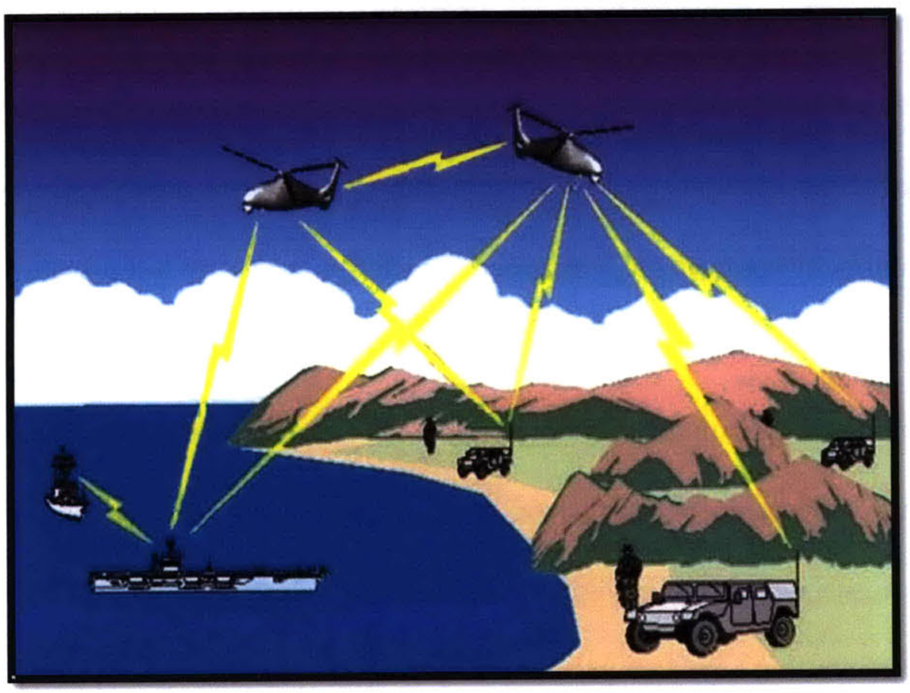

Figure 1: Coordinated Operations with Networked UxVs [2]

Automation is of utmost importance since computers provide the technological capability of quickly analyzing and editing a mission plan while accounting for every known mission constraint and requirement. However, computer optimization algorithms are"brittle" since they only account for quantifiable variables coded in the design of the system [3]. As a result, human judgment is an imperative part of the human-machine system. In highly autonomous systems, humans must rise to the role of human supervisory controllers. "Supervisory control means that one or more human operators are intermittently programming and continually receiving information from a computer that itself closes an autonomous control loop through artificial effectors and sensors to the controlled process or task environment" [4].

Automation is designed to lower the operator's information processing demands in order to improve situational awareness and increase performance. However, an approach involving high levels of automation can be counterproductive [5]. As automation directly controls the unmanned vehicles, humans can fall prey to "the ironies and paradoxes of automation" [6]. It is said that the more reliable the automation, the worse human operators perform in the monitoring task [6]. Increased automation can lower an operator's workload too much, leading to mental underload, which can cause a decrement in vigilance, or sustained alertness, and lead to boredom. It has been shown that boredom produces negative effects on morale, performance, and quality of work [7]. Unfortunately, as increased automation shifts controllers into system management positions, loss of vigilance, monotony, and boredom are likely to proliferate [8].

1.2 Operational Benefit

Although today's military employs a team of people to operate a single UxV, advances in automation technology seek to invert the ratio of operators to UxVs so that, in the future, one human operator will be able to control multiplue UxVs [9]. The vision is to have a single operator controlling land, air, and sea vehicles of all different types from the same supervisory control interface. As human supervisory control of UxVs becomes more prevalent, networks of vehicles equipped with collaborative autonomy will become reality [10]. This research hopes to provide future system designers with an assessment of the impact that low workload has on supervisory control of multiple UxVs.

To this end, a long duration, low workload study was conducted using a multiple UxV simulation. This human supervisory control experiment involved a search, track, and destroy mission scenario. The mission was designed to be a realistic situation with a dynamic environment full of moving emergent targets, including some hostiles. The simulation specifically involved a high level of automation in order to induce boredom. This simulation mimics real world Unmanned Aerial Vehicle (UAV) missions, which involve low workload and range from 8 to 12 hours.

In addition to providing research support for future multi-UxV objectives, this study applies to a myriad of domains where boredom is prevalent in current operations. For instance, UAV Predator pilots face vigilance and boredom issues due to long duration, low workload missions. The aviation world also suffers from these problems, as in the Northwest Airlines incident of 2009 where the pilots overshot their destination by 150 miles due to loss of vigilance and situational awareness [11]. This research also applies to scenarios such as air traffic control in low traffic situations, transportation system monitoring, and process control supervision. Already, the prevalence of human-machine systems has caused increased interest in vigilance research [12].

This long duration, boredom research in the context of networked UxVs is invaluable because, despite the growing need for boredom and vigilance research [13], there is a shortage of research on this topic [14, 15]. The occurrences of vigilance degradation and boredom are not well understood, and neither are their outcomes [16]. Literature reviews on these topics are outdated [17, 18]. In light of current technological advances and the necessity of boredom research on vigilance tasks, it is even more important to update research on this topic.

1.3 Thesis Organization

Chapter 1, Introduction, outlines the motivation and operational benefit for this research.

Chapter 2, Background, provides information on workload, vigilance, boredom, and fatigue, and their implications on unmanned vehicle operations. It also details the research questions and hypotheses of this thesis.

Chapter 3, Experimental Evaluation, describes the procedures and design of the Low Taskload, human-performance experiment used to test the hypotheses of this research.

Chapter 4, Results and Discussion, presents the results of the analysis for each research question immediately followed by discussion.

Chapter 5, Conclusion, states the findings of this study and provides recommendations for future work.

2 Background

The literature review presented in this chapter is the structure supporting the experimental methodology of this research. The three pillars of this research are workload, vigilance, and boredom. This chapter explains the theory behind low workload with regard to performance and discusses the vigilance decrement associated with low workload. Empirical evidence for measuring vigilance is presented, followed by empirical evidence for measuring boredom. Pitfalls of boredom and fatigue are discussed in the context of current unmanned aerial vehicle domains. Furthermore, this chapter sets the stage for the experimental testbed used in this study by describing a previous single-operator UxV experiment on moderate-level workload and performance. This chapter culminates in the presentation of the five research questions and hypotheses investigated.

2.1 Workload

Workload plays a pivotal role in the performance of a human-automation system. Workload is an individual's perceived level of busyness, while taskload is the amount of work imposed upon an operator [19]. Workload and taskload often go hand in hand; however, a person who is easily overwhelmed may perceive a moderate taskload as high workload. The Yerkes-Dodson law, which explains the link between workload and performance, is now discussed as a motivating factor for this research to determine whether performance, in fact, declines in a parabolic fashion as workload decreases.

2.1.1 Yerkes-Dodson Law

The Yerkes-Dodson law describes the relationship between workload and performance as shown in Figure 2. The Yerkes-Dodson "law" nominally depicts a drop in operator performance when the operator is over-worked or under-worked.

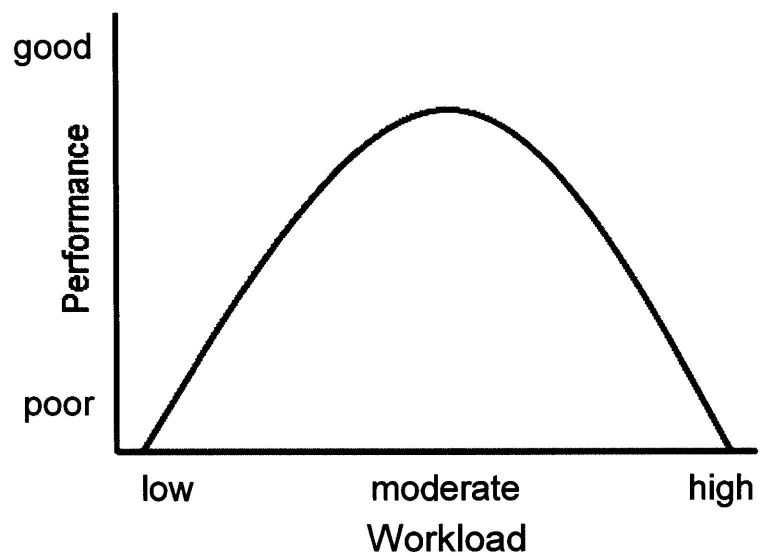

Figure 2: Graphical representation of the Yerkes-Dodson Law

Although the Yerkes-Dodson law, created in the year 1908, originally related arousal to performance [20], the law has been extended to incorporate workload in the place of arousal [21, 22]. A relationship similar to the Yerkes-Dodson curve suggests that the drop in operator performance during low arousal is due to human complacency, while the drop in performance during high arousal is a result of overload [23].

Research shows that operators controlling multiple UxVs perform significantly worse under high operational tempos [24] [25] [26]. A metric that objectively describes an operator's workload is utilization, or percent busy time. It has been shown that performance significantly degrades when supervisory control operators are tasked beyond 70% utilization [24] [27] [28]. Although a general consensus recognizes that performance drops off according to the Yerkes-Dodson law at high levels of workload, little is known about whether the low end of workload actually mirrors the same plummet in performance, particularly in the context of supervisory control of multiple UxVs in a highly autonomous system.

The Yerkes-Dodson law is notional, and steep drops in performance have only been reported for high workload [24]. It has been argued that the Yerkes-Dodson measure of workload, or arousal, is lacking in three areas: predictive capability, clarity, and unitary construct [22, 29]. Thus, the Yerkes-Dodson curve has serious drawbacks for predicting performance. Several sources claim that the connection between workload and performance is much more complex than an over-simplified, inverted "u-shape" curve suggests [30-32].

This thesis research seeks to determine the validity of the Yerkes-Dodson relationship between performance and low workload. A long duration, low workload experiment using a networked UxV supervisory control simulation was conducted to measure performance among three groups of varying taskload. This experiment was designed to compare performance across three low levels of workload, assuming that taskload corresponds with workload.

2.2 Vigilance

Vigilance is denoted as a state of being alertly watchful, especially to avoid danger, and is often required in a military, supervisory control context. The human tasks of monitoring and decision making for a networked UxV system can be considered controlled processes, which are described as serial tasks requiring effort under an individual's direct control [33]. It is known that vigilance decrement is an inherent part of controlled processing [34]. Some researchers refer to vigilance decrement as a decrease in attentional capacity, which is a result of overload from high mental workload [35-37]. However, other researchers state that vigilance decrement is caused by attentional withdrawal from low workload [38-40]. This research focuses on vigilance associated with low workload.

2.2.1 Measuring Vigilance

Measuring vigilance may include objective, physiological, and subjective instruments [41]. Vigilance can typically be measured objectively according to four manifestations of how quickly people can detect critical events: (1) *target detection* rate, or *hit* rate, (2) *non-target detection* rate, or *correct rejection* rate, (3) *failure to detect targets* rate, or *omission* rate, and (4) *incorrect identification of non-targets as targets* rate, or *false alarm* rate [42]. UxV operations of the future, which include highly autonomous systems, will require sustained vigilance due to the need for prolonged monitoring and persistent surveillance. Vigilance research suggests that a performance trade-off exists between active and passive sustained monitoring [42].

For example, on such study involved a passive, sonar target detection environment with target tones sounding in a noise background at a mean rate of 10 per minute, and irrelevant probe tones playing at intervals of 2 to 4 seconds [43]. Participants listening for sonar target tones were asked to make false detections of irrelevant probes. During the 28-minute test session, the participants' response rates fluctuated for minutes at a time, indicating a long-term change in performance. Response rates of the false detections declined after only 2 to 3 minutes of task performance, and subsequent response rates stayed below 70 to 80% of initial rates. According to the study, it was shown that averaged false detections of the frequent, irrelevant probe tones provide an accurate estimate of alertness level. However, measuring detection frequency and accuracy is not the best representation of vigilance.

Nevertheless, similar studies measure vigilance using operator detection times. Two studies on air traffic control (ATC) en route monitoring determined that the time to detect conflicts and the frequency of missed traffic conflicts increased significantly over the course of just two hours [44, 45]. This degradation in vigilance over a 2-hour period justifies the need to perform studies with even longer vigilance tasks. For example, the average shift length of a UAV pilot is 12 hours for the US Air Force and 8 hours for the US Army.

Cerebral blood flow has been linked to vigilance performance. When parts of the brain become metabolically active, the by-product of mental exertion, carbon dioxide (CO_2), increases [46]. The human body subsequently reacts by speeding up the blood flow in that area to remove the waste gas. A previous Transcranial Doppler sonography study showed that cerebral blood flow velocity significantly declined linearly over time as participants performed vigilance tasks involving signal detections in the auditory and visual realms [46]. In addition, participants experienced a general reduction of responsiveness in vigilance tasks during four 10-minute tests. The decline in vigilance and cerebral blood flow suggests that information processing resources are not replenished as quickly as they are consumed over long periods of time.

Similarly, it has been shown that the electroencephalographic (EEG) power spectrum changes accompany minute to minute fluctuations in alertness [47]. Fifteen subjects participated in a dual-task simulation of visual and auditory sonar target detection. Each subject performed three 28-minute sessions. Accurate, non-invasive, nearly real-time estimates of an operator's global vigilance were measured with EEG recorded from only two central scalp sites. Data from sessions where at least 25 lapses in target detection were recorded was compared against EEG measurements. Power spectra were sorted by local error rate, and EEG power was correlated with changes in error rate. The results showed that a monotonic relationship exists between minute-scale changes in performance and the EEG spectrum. This research showed that changes in alertness can be measured by EEG power spectrum changes.

Although vigilance has been measured using detection rates and physiological signals, it has been suggested that most vigilance studies have been conducted in strict laboratory environments with far more stimulus events than are realistic [48]. Instead, the number of concurrent operator tasks needs to be minimized for researchers to discover subtle changes in operator behavior [48]; that is, the experimental setting needs to promote boredom. Others have noted that measuring vigilance in low workload experiments is actually linked to boredom measurement [49]. Rather than measuring vigilance through response times and physiological recordings, this research focuses on measuring vigilance through performance-based and attention-based measures of boredom, discussed next.

2.3 Boredom

Boredom can be a major problem in the supervisory control setting because people become under-stimulated to the point where sustaining mental effort is impossible. There is evidence to suggest that task underload results in operator performance degradation [50]. It has been suggested that boredom encompasses two components: cognitive and affective [51]. The cognitive component comes from an operator's perception of the task at hand. If the task seems unimportant or nonchallenging, the operator becomes cognitively disinterested. The affective component of boredom relates to the operator's emotional perception. Feelings of frustration, dissatisfaction, melancholy, and distraction represent the affective side of boredom [51]. The following subsections describe the impact that boredom has on operator performance in human supervisory control tasks and present methods for identifying boredom. Additionally, boredom proneness as it relates to crew selection and the unmanned aerial vehicle domain is discussed.

2.3.1 Measurable Performance Impact of Boredom

Performance degradation can be measured as a function of boredom. Air traffic controllers in low taskload environments, such as en route monitoring of aircraft, can be susceptible to boredom, unlike the busy terminal operators. Studies on ATC monitoring tasks showed that participants reporting high boredom were more likely to have slower reaction time and worse performance than participants reporting low boredom [52] [53]. Similarly, participants who reported higher subjective, task-related boredom also had slower reaction times. People recognize when they are bored, as shown by the participants' boredom reports matching their slow reaction times.

Furthermore, a study of American air traffic controllers showed that a high percentage of system errors due to controller planning judgments or attention lapses occurred under low traffic complexity conditions [54]. Consequently, system designers need to make an effort to prevent boredom and avoid complacency of controllers in order to sustain vigilance in low workload conditions [55].

Specific factors influencing boredom and monotony have been examined in the context of ATC. It has been suggested that task characteristics (e.g. repetitiveness, traffic density) may interact with individual influence (e.g. personality, experience, age) and work environment in a way that causes monotony and boredom [16]. This research was a first step in examining monotony from a perspective of individual factors in the hopes of guiding crew selection, training, and understanding of how individual factors affect critical states [16]. In the same way, the research of this thesis seeks to identify participants' characteristics that influence boredom in a low workload environment.

2.3.2 Identifying and Measuring Boredom

People show expressions through channels of communication, such as body language, facial expressions, tone of voice, and posture, to name a few. Characterizing and recognizing the human emotion of boredom is essential for diagnosing workload issues in the context of futuristic UxV operations. In a previous study, a three-dimensional optical flow tracking system was used to rate participants' boredom levels as they watched a stream of boring videos [56]. Two judges watched footage of participants watching these boring videos. The judges watched videos of a participant's head and shoulders, and had two screens of footage showing the participant's left and right sides, respectively. The two judges identified events as a team, and then individually rated whether the event showed any change in attention state. The judges' boredom ratings were analyzed in conjunction with head position data to objectively identify boredom events. A similar video coding methodology was used in another study [57]. Slumping posture from the head position data in conjunction with judges ratings of boredom from the participants' facial expressions indicated when boredom was occurring.

Video coding shows that humans deal with boredom in different ways. Some individuals are more prone to boredom than others. Personality, attention span, and personal interests can affect whether people become bored easily. A study showed that subjects with low boredom proneness outperformed high boredom prone subjects and reported less boredom in a flicker detection vigil [58]. Taking into account boredom proneness could improve crew selection of monitoring tasks.

2.3.3 Boredom in Unmanned Aerial Vehicle Domains

Persistent surveillance is accompanied by persistent, boredom-inducing tasks. Boredom is prevalent in unmanned aerial vehicle operations, amid rare and short moments of critical, hostile situations. An ex-A-10 pilot flying Predators is

> "likely to seek out action, for example, by monitoring the banter on the secure chat rooms used by commanders to communicate in battle. 'Highly skilled, highly trained people can only eat so many peanut M&Ms or Doritos or whatnot,' he said. 'There's the 10 percent when it goes hot,

when you need to shoot to take out a high-value target. And there's the 90 percent of the time that's sheer boredom—12 hours sitting on a house trying to stay awake until someone walks out [59].'"

2.3.4 Fatigue

Fatigue impacts long duration missions, manifested as a lack of mental alertness, regardless of the level of workload being expended throughout the mission. Fatigue, like boredom, becomes a primary problem in supervisory control of multiple unmanned vehicles. Fundamentally, fatigue is driven by a chronic lack of sleep. However, a relationship exists between boredom and fatigue.

In a Predator operations study, "graphical analysis of subjective boredom ratings found 92 percent of pilots reported 'moderate' to 'total' boredom" [60]. It is interesting that a study focused on researching fatigue also showed high ratings of subjective boredom. The boredom caused slower responsiveness, which resulted in problems with performance and crewmember morale.

Merely limiting flying time of shift workers proved to be a poor safeguard against fatigue. Even a four-hour work shift still resulted in fatigue and boredom [60]. The harmful effects of fatigue and boredom must be investigated before futuristic, highly-automated operations of multi-UxV control become reality.

2.4 Empirical Evidence for Possible UxV Vigilance Problems

A previous study that attempted to examine the impact of moderate workload in supervisory control of multiple UxVs yielded unexpected results that suggest vigilance and boredom could be significant factors in such an environment. This experiment was conducted using the Onboard Planning System for Unmanned vehicles Supporting Expeditionary Reconnaissance and Surveillance (OPS-USERS) test bed [61]. The simulation allowed a single operator to supervise multiple autonomous UxVs in a search, track, and destroy mission. The operator was assisted by an automated planner for scheduling the UxVs' search, track, and destroy tasks. In addition, a decision support tool allowed the operator to alter automation-driven schedules and approve desired plans. As will be discussed in detail, even in a moderate workload study, there was evidence to suggest that vigilance could be a problem in supervisory control of multi-UxVs.

The objective of the operator was to command multiple, heterogeneous UxVs for the purpose of searching the area of responsibility for hidden targets, tracking targets, and approving weapons launches [26]. The UxVs in this experiment included two rotary-wing Unmanned Aerial Vehicles (UAVs), one Unmanned Surface Vehicle (USV), and a Weaponized Unmanned Aerial Vehicle (WUAV). Once a target was found, the user designated the target as hostile, unknown, or friendly, and assigned it a priority level. One or more UxVs continually revisited hostile targets to track their positions until the WUAV was able to destroy the hostiles. Operators were required to approve all weapon launches from the WUAV. Unknown targets were also revisited as often as possible, tracking the targets' movements. Provided with intelligence via a chat box, the operator could re-designate unknown targets as hostiles or friendlies. The operators could create search tasks, given unsearched locations on the map, for UxVs to explore. The operators spent much of the mission time monitoring the system, while the auto-planner prompted replanning sessions for re-evaluating the unassigned tasks that needed to be scheduled.

2.4.1 Experimental Apparatus

The interface details can be found in Appendix A. Figure 3 shows the top layer display of the human-computer interface (HCI) that was used for this study. This top layer display, known as the Map Display, shows symbols representing the UxVs, search tasks, loiter tasks, and targets.

A birds-eye view of the mission area is shown with representational symbols of UxVs, targets, tasks, etc. The symbols correspond with Military Standard 2525 [62]. These symbols include: UxV symbols that represent the four vehicles moving over the map; search task symbols, which are markers on the map that represent an operator-designated location for the UxVs to explore in search of hidden targets; target symbols such as hostiles, unknown targets, and friendlies found roaming the map that are to be tracked; and loiter symbols, or points on the map for the weaponized vehicle to

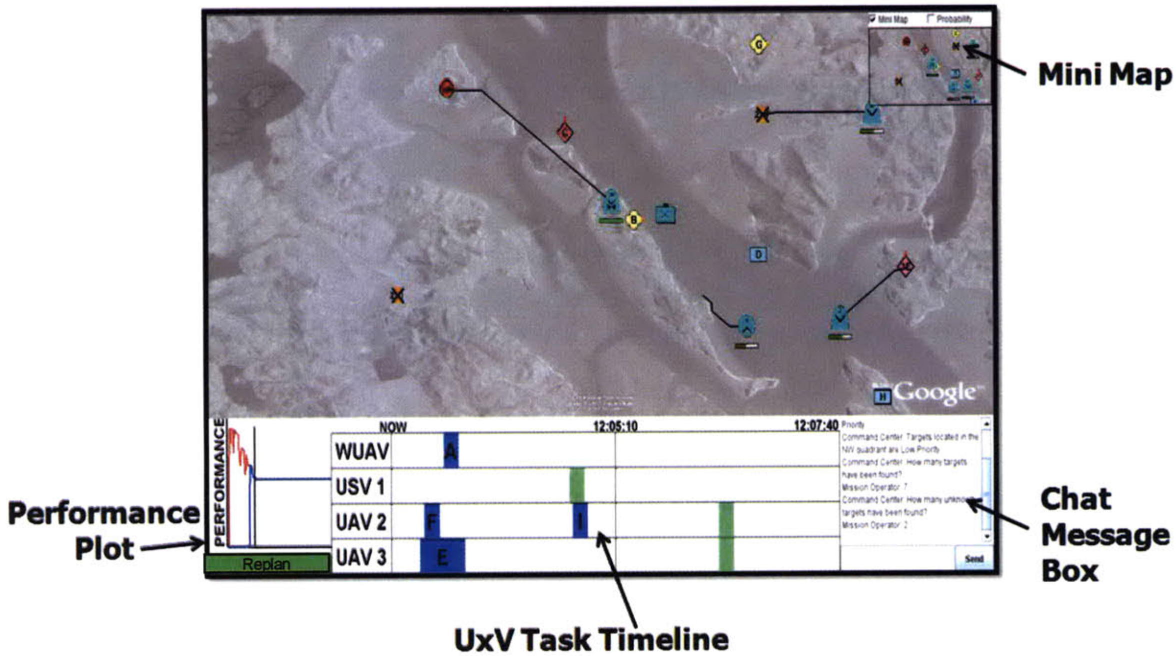

Figure 3: Map Display

hover over while waiting to destroy the next hostile target. The upper right-hand corner of the Map Display is equipped with a mini map that shows the symbols for UxVs, search and loiter tasks, and targets as they appear on the map. Since the Map Display can be zoomed in, it is convenient to glance at the mini map for a quick view of the overall picture. This feature can be turned off by un-checking the mini map box above the mini map itself.

The UxV timeline at the bottom of the Map Display gives temporal event information for the next five minutes into the future, indicated in military time. Green bars in the interface indicate times of refueling, and blue bars indicate times of performing a task. The letter of the task (whether a search task or target-tracking task) appears in the blue bar. White space indicates vehicle idle time or travel time between tasks. The timeline moves to the left as time progresses.

The lower left-hand corner of the Map Display portrays a performance plot, shown in Figure 4. The automation analyzes the current schedule, predicts mission performance by the end of the mission time, and calculates a score. The score is calculated based on a non-dimensional cost function that accounts for task priority and completion, target tracking, hostile target destruction, and coverage area. The red score represents the automation's predicted score. The blue score represents the actual score attained by the human-automation system. When the predicted score surpasses the

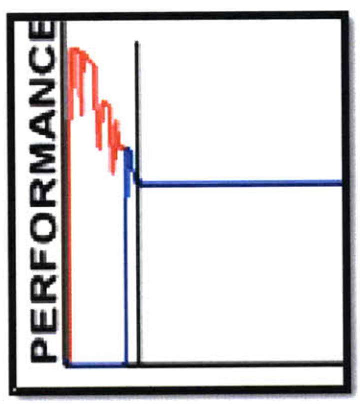

Figure 4: Performance Plot

actual score, the auto-planner is proposing that better performance could be achieved if the operator accepts the proposed plan. On the other hand, when the actual performance exceeds the predicted curve, the human operator has changed the tasking in a way that results in better system performance than the automation predicted [63]. The performance plot moves to the right as the score changes over time.

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 chat message box shown in Figure 5 gives important information dictating priority levels for targets. Chat messages are accompanied by an auditory tone common to modern-day instant messaging programs. In addition, the chat box outline blinks until the operator acknowledges the received message by clicking in the chat box. Sometimes chat messages require responses to questions, such as, "How many targets have been found?" The operator must type the answer in the message input window and click "send."

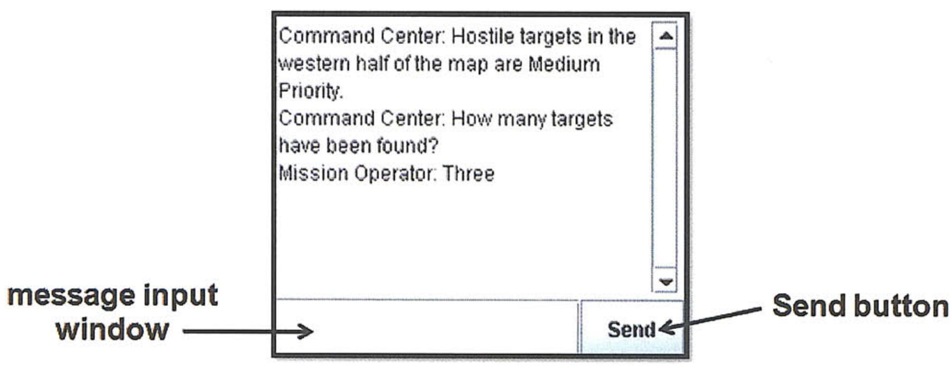

Figure 5: Chat Message Box

2.4.2 Operator Tasks

The main tasks for the operator include: creating/editing/deleting search tasks, identifying targets, replanning, and destroying hostile targets.

2.4.2.1 Search

A primary mission objective is to search uncharted territory. The UxVs automatically search the area of interest using their own onboard computer search algorithm, which is an A* search method. However, it has been shown that systems with human operators are better than purely automated systems at ensuring the entire map area is covered in the search [26, 63]. The operator can create a search task at a particular location by right clicking the location on the map, which brings up the search task creation window, shown in Figure 6. The operator designates the priority level and temporal requirements of the search task. The operator can also create loiter tasks using the search task creation window. Right clicking an existing search task allows the operator to edit using the same window.

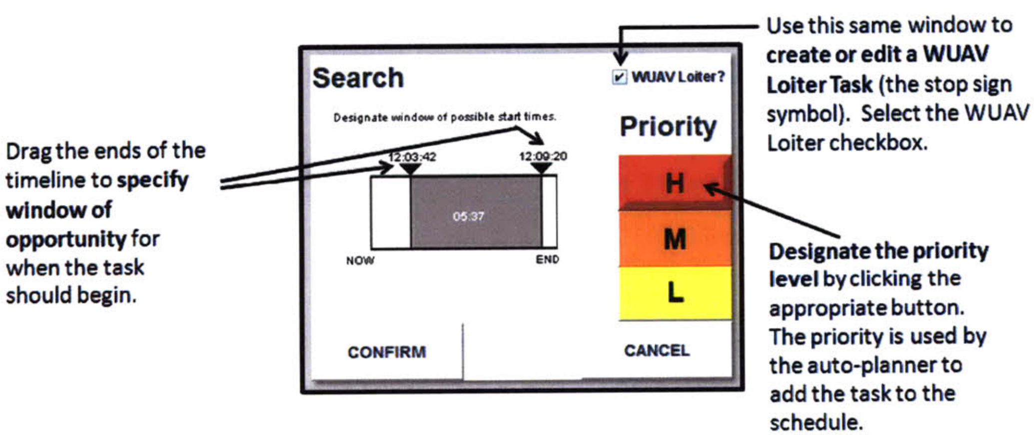

Figure 6: Search Task Creation Window

2.4.2.2 Identify Targets

The UxVs have automatic target detection capability in the futuristic scenario of the OPS-USERS simulation. The target identification window pops up automatically when one of the UxVs discovers a target. For experimental purposes, the target identification task was simplified to recognizing the target symbols rather than analyzing actual imagery. The operator must pan through the target identification window until the target symbol becomes visible. The operator then classifies the target symbol as hostile, unknown, or friendly and designates a priority level of high, medium, or low priority using intelligence information from the chat message box. Figure 7 shows the sequence of target identification.

Figure 7: Target Identification Window Sequence

2.4.2.3 Approve Weapons Launch

When a target is identified as hostile, it must be destroyed by the WUAV while being tracked by the UxV that found it. Operator approval must be given before the WUAV is allowed to destroy a hostile target. The missile launch approval window shown in Figure 8 pops up automatically when the WUAV sights the hostile target for destruction.

Figure 8: Missile Launch Approval Window

The operator must pan the screen for a direct view of the target and click the red "approve launch" button to destroy the target.

2.4.2.4 Replan

The automation prompts the operator to replan by approving new UxV schedules. However, the operator can also initiate the replanning. Given the current schedule, the automation's proposal, and potentially changing mission priorities, the operator can change UxV schedules via the replan display. The replan display is a decision support tool known as the Schedule Comparison Tool (SCT), shown in Figure 9. The green "replan" button at the bottom left corner of the Map Display shown in Figure 3 allows the user to view the SCT.

All mission objectives, including search tasks and targets to be tracked and/or destroyed, are either assigned or unassigned via the SCT. The gray areas around the black "assign" triangle in the SCT display the tasks not yet assigned to any UxVs. Operators are able to click and drag unassigned objectives into the central "assign" area, essentially querying the automation about whether the particular objective can be assigned. Sometimes not all tasks can be assigned. Subsequently, the new assignment of a task can cause other tasks to become unassigned. Tasks that can no longer be assigned pop out of the black "assign" area and move to the gray area of unassigned tasks.

The three geometrical forms at the top of the SCT are configural displays and show three schedules. The dark gray form on the left is the current schedule being carried out by the UxVs. The green form on the right is the newest proposed schedule from the automated planner. The blue schedule in the center is the working schedule that results from the user querying the automation to assign particular tasks. Thus, the proposed schedule represents a highly automated solution; the working schedule promotes a more collaborative effort between the human and computer, which has been shown to improve operator performance and situational awareness in similar complex settings [65-67].

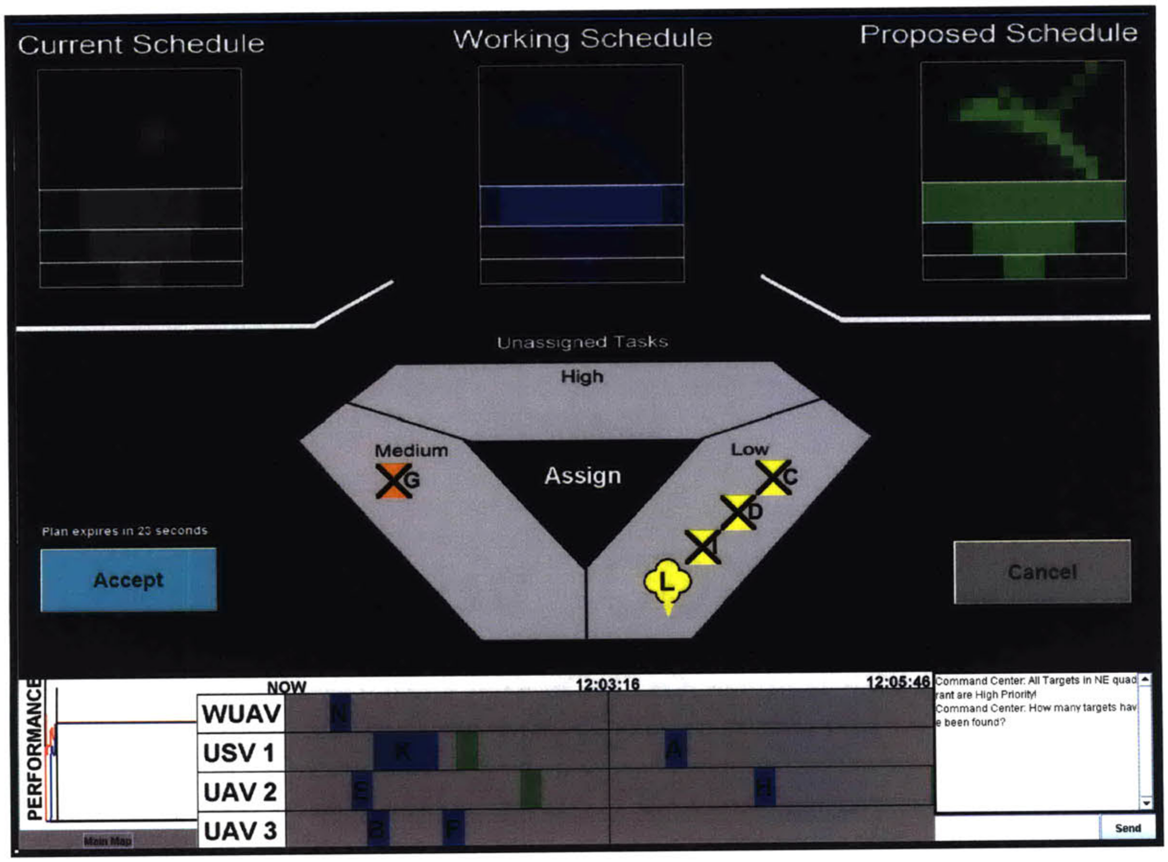

Figure 9: Schedule Comparison Tool

Each configural display is composed of two parts: an upper rectangle and a lower rectangle separated into three bars. The configural display is shown in Figure 10. The top rectangle represents the map area that will be covered for a given schedule. The more colorful the area, the better searched the map will become using that schedule. The bottom hierarchy of bars shows the percentages of high, medium, and low priority tasks to be completed for a given schedule. The more color-filled a bar appears, the more of that task priority is being done. When a task is assigned, the corresponding bar changes shape with a ghosting effect in order to visually draw attention to what has changed. The white overlay shown in the high priority bar of Figure 9 is the result of the ghosting effect. This white overlay depicts the previously smaller percentage of high priority tasks being assigned.

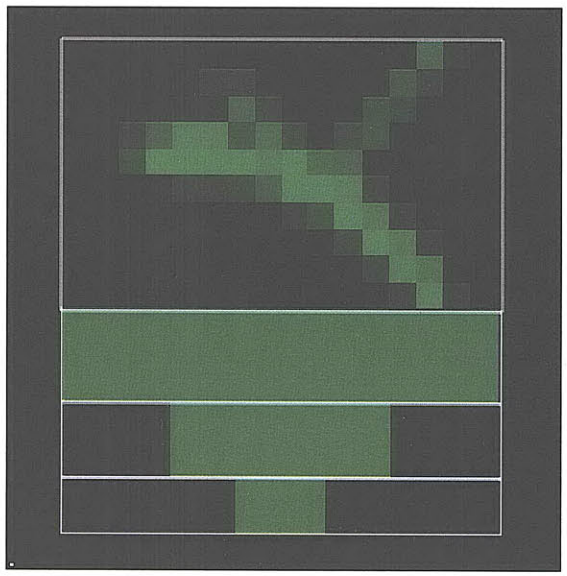

Figure 10: Configural Display

2.4.3 Moderate Workload Experimental Results

The original study of moderate workload on the OPS-USERS testbed yielded interesting results that motivated this research on low workload. The moderate workload replan interval experiment assessed operator workload and performance in three automation-generated replan intervals. Specifically, the rate at which the operator was required to collaborate with the automation using the SCT was modulated over three experimental trials. The intervals for replanning were 30 seconds, 45 seconds, and 120 seconds. The order was counterbalanced across the test sessions for thirty-three participants [26].

This study showed that people performed better when they worked with the automation's prescribed replanning rates, rather than ignoring the automation and operating under their own discretion for when to replan [63]. The interesting result from this experiment that motivated this thesis research is that even though participants who consistently responded to the automation's replan prompts, deemed consenters, were unable to maintain the automation's prompted replanning rate at the lowest interval of 120 seconds; that is, the consenters of the experiment replanned more often when the automated replan interval fell below a comfortable threshold of workload. This finding shows that humans have difficulty maintaining low levels of workload, and further research was needed for the low workload scenario of this simulation.

2.5 Research Questions

The five research questions this thesis seeks to answer aim to explore different facets of the overall question: how do people behave under long duration boredom? This study is retrospective in nature, and these research questions are provided in order to approach the research in a specific, measurable way. A hypothesis was devised for each research question to help guide the analysis, but not limit it. These questions are as follows:

1. **Does the Yerkes-Dodson curve hold true for low workload?**

It has already been shown that high workload does cause performance to plummet, and the Yerkes-Dodson curve is valid for high workload conditions [24] [25] [26]. However, the amount of research conducted in long duration, low workload environments for human supervisory control is small [14, 15]. Persistent surveillance tasks and sustained monitoring tasks are common in human supervisory control settings. With these jobs becoming ever more prevalent as automation increases, the effect of sustained low workload on performance needs to be understood [13]. Does low workload really cause performance in supervisory control to plummet as the Yerkes-Dodson curve suggests? It is hypothesized that low workload data from this experiment will show that the Yerkes-Dodson law is not correct for low workload.

2. **How does low taskload affect operator utilization?**

This research seeks to identify how participants react to low system requirements of taskload. Will participants become disinterested and let their interactions with the interface fall below the required amount to perform tasks? Will participants overindulge in interacting with the system in order to stay alert? In this study, participants have the freedom to interact with the system as much or little as they please. This experiment is a unique opportunity to learn about human nature by studying human-system interaction levels under low workload conditions. It is hypothesized that taskload, modulated by replan interval in this low workload study, will have a positive relationship with utilization, or total interactions with the system.

3. **How does the low workload environment affect operator attention?**

Knowing how low workload affects performance and utilization is not enough. Understanding attention allocation is key to discovering the toll that sustained low workload takes on human operators. It is hypothesized that operators will spend most of their time in divided attention (coping with boredom by multitasking), some of their time completely distracted (due to boredom), and the least amount of their time in directed attention (because of low workload and disinterest).

4. **Can performance be predicted in a low workload environment?**

Being able to predict performance in persistent surveillance tasks could be a tremendous benefit to the supervisory control domain. Predicting performance could lead to preventing vigilance decrements and fatal errors before they happen. In order to predict performance, attention allocation as it relates to utilization and performance will be investigated. Perhaps performance can be predicted knowing how focused a person is apt to be. It is hypothesized that operators with higher percentages of directed attention will perform better, as predicted by statistical models.

5. **Does vigilance decrease over time?**

Vigilance decrements are often associated with long duration, supervisory control tasks. But does this phenomenon really occur? It is important to research what really happens to an operator's sustained alertness in the context of a multiple UxV mission setting, since this scenario is the future of unmanned vehicle operations. The literature review revealed that current measures of vigilance create an unrealistic testing environment, and a boredom study is needed to discover subtle changes in behavior and effectively assess vigilance [48]. Accordingly, it is safe to assume that vigilance can be measured by attention state changes from hour to hour. It is predicted that operators' amounts of directed attention per hour will decrease with each subsequent hour. It is hypothesized that, in this way, vigilance will decrease over time.

These research questions stem from the three pillars of the background presented in this chapter: workload, vigilance, and boredom. The following chapter describes the methodology for answering these research questions and creating an overall assessment of the impact that low workload has on supervisory control of networked unmanned vehicles.

3 Methodology

This chapter describes the long duration, low taskload human performance experiment used to test the research hypotheses detailed in the previous chapter. Thirteen groups of 3 participants endured a 4-hour experimental session acting as independent operators engaged in supervisory control of networked UxVs. The simulation was a search, track, and destroy mission conducted on the OPS-USERS testbed detailed in Chapter 2. This chapter discusses participant information, the apparatus, testing procedures, and experimental design.

3.1 Participants

Thirty-nine participants were tested 3 at a time. Complete test data was collected for 30 participants, which included 11 females and 19 males. Data from 9 of the participants was incomplete or unusable because of system software failures. Forty-three percent of the participants had military experience. Participant age ranged from 19 to 32 with a mean of 23 years of age and a standard deviation of 3 years; this age range is typical of current unmanned vehicle operators in the military. Each participant was classified as either a "gamer" or "non-gamer" based on their video gaming experience revealed in the demographic survey. Participants who played games more than once a week were considered gamers. Each participant signed a consent form, shown in Appendix B.

Sixteen of the 39 participants originally participated in the moderate workload study discussed in Chapter 2. The remaining 23 participants received equivalent training on the moderate workload testbed. New participants learned about the interface via the self-paced tutorial used for the moderate workload experiment and participated in a mock-experiment on the moderate workload testbed for a total of approximately 1.5 hours. This training was performed to ensure consistency of practice among all participants for this long duration, low workload study.

3.2 Apparatus

This section focuses on the modifications made to the OPS-USERS system for converting it to a long duration, low taskload scenario. The test session for this experiment lasted 4 hours, as opposed to the 10-minute session in the moderate workload study [26]. Each participant only performed one 4-hour test session for a given replan interval. Each operator workstation included two 17-inch Dell TFT LCD monitors connected to a Dell Dimension tower containing a Pentium D 2.80GHz CPU and 2.00 GB ram. The interface was displayed on the left monitor with the right monitor being open for participant prerogative use.

To make the workload lower than the moderate workload study, the unmanned vehicles moved 10 times more slowly across the map. It took almost an hour for a vehicle to move from one side of the map to the other, which appeared extremely slow since it only took a couple of minutes for a vehicle to traverse the map in the moderate workload study. The scenario also had only 4 hidden targets to find in the 4-hour mission, unlike the 10 targets in the ten-minute moderate workload scenario. Moreover, the participants were prompted to replan only once every 10 minutes, 20 minutes, or 30 minutes, depending on their issued replan interval, as opposed to every 30 seconds, 45 seconds, or 120 seconds in the moderate workload scenario. All of these modifications to target number, vehicle speed, and replan interval were done in an effort to center the participants' workload around 10% utilization, unlike the 70% utilization goal in the moderate workload scenario. The target utilizations for the three replan interval groups were 15%, 10%, and 5%.

An additional way of maintaining low operator taskload throughout the entire session was to ensure that the 4 targets could not be found all at once. One of the 4 targets was "uncloaked" at the beginning of each hour. Thus, 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. This "uncloaking" activity ensured consistently low workload for operators throughout the 4-hour study. The participants were unaware that targets remained hidden and only emerged later in the simulation.

3.3 Experimental Procedure

The 4-hour, low workload test session was prefaced by pre-experiment paperwork, including consent forms, demographic and training surveys. Participants were tested 3 at a time, but each performed separate simulations. Participants were knowingly videotaped during the test session to capture behaviors exhibited throughout the study, as shown in Figure 11. Workload and performance metrics were collected automatically by the simulation without interrupting the participants.

3.3.1 Paperwork and Practice

Participants completed a demographic survey, which can be found in Appendix C. Details about the demographic results can be found in Appendix D. After completing the paperwork prior to the experiment, participants completed a self-paced, refresher tutorial and were allowed to ask questions. Following the self-paced refresher tutorial, all three participants completed an interactive practice session during which they practiced all of the tasks that would be required during the four-hour test session. Participants could practice as long as needed to feel comfortable with the interface, usually about 10 minutes. After practicing, each participant filled out an exit form that illustrated his or her confidence level in understanding the interface and mission scenario. The exit survey for interface understanding can be found in Appendix E. The overwhelming majority of participants answered "confident" or "very confident" (with only 6 of 39 feeling "somewhat confident") and indicated they understood the interface functionalities. Any problem areas were covered again. After all questions were answered, the test session commenced.

3.3.2 Test Session

Three participants were tested at a time in a mock command and control center shown in Figure 11. All operators' scenarios were independent of one another; i.e., there was no need or opportunity for collaboration designed into the scenarios. Because of the long duration of the study, three participants were tested at a time, both to reduce overall experiment time and to provide possible sources of distraction. Unmanned vehicle operating environments typically contain multiple personnel who are often responsible for dissimilar tasks, so this environment was representative of typical command and control centers. Each participant assumed supervisory control of their own set of 4 heterogeneous, unmanned vehicles in their own territory.

Figure 11: Three Subjects in the Test Room

Participants assumed limited control over the UxVs, assigning search and target-tracking tasks to the system network but not to particular UxVs. This lack of direct control was identical to that of the moderate workload experiment discussed in Chapter 2. Participants employed a weaponized unmanned aerial vehicle to destroy hostile targets. The underlying automation concurrently analyzed the mission as it progressed and proposed new plans at predetermined intervals. The participants viewed these proposals via the SCT interface shown in Figure 9, which allowed them to edit and accept the new plans.

Potentially distracting material was available to the participants during the experiment, such as internet access via one of the workstation interfaces that was not in use, magazines, refreshments, anything the participants had with them in their backpacks (including cell phones or books), and each other. Refreshments were provided to the participants, and the same food varieties were served to all participants. Participants could bring their own lunches if they so desired. Each set of 3 participants was left alone in the mock command and control room during the study. However, participants were knowingly videotaped for the duration of the study. In addition, screen capture software was used to record the interface interactions. The test administrators remained in an adjacent room and came into the test room 4 separate times to check on the participants throughout the study duration. During the experiment, participants could leave the test environment to go to the restroom at any time; in this case, the test administrator paused the simulation in order to preserve the participant's data. Upon return, the experiment administrator informed the participant that the scenario remained stable and uneventful during his or her absence, and the participant resumed the simulation. Participants only left the room to go the restroom. Thirty minutes prior end of the simulation, the timeline grayed-out, indicating that no future events were visible as the simulation came to a close.

Following the test session, participants filled out a survey, where they indicated how busy they felt, their confidence in the actions they took, and how well they felt they performed. The post-experiment survey can be found in Appendix F. They also indicated whether they were distracted or not, and listed any distractions they encountered during the test session. Participants were compensated $125 for their efforts and were also eligible to win a $250 Best Buy gift card for the best performance.

3.4 Experimental Design

This long duration, low taskload simulation was designed to investigate low workload as it relates to operator performance. Taskload was controlled by simulation-prompted events that required major decision making. The experiment was originally designed to be statistically evaluated using a One-way Analysis of Variance (ANOVA) model with 3 factor levels represented by the 3 replan intervals.

3.4.1 Independent Variable

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 in schedule decision making. Each participant was given a fixed replan interval of either 10 minutes, 20 minutes, or 30 minutes; these replan intervals were intended to induce utilization levels of 15%, 10%, and 5%, respectively. This prediction was estimated based on the previous study and pilot testing of the low taskload scenario.

3.4.2 Dependent Variables

The dependent variables include objective workload, objective performance metrics, subjective self-rated performance metrics, and attention state metrics obtained via video data.

3.4.2.1 Workload Metrics: Utilization

Utilization, or percent busy time, has been used to detect subtle changes in workload during time-pressured scenarios, similar to this OPS-USERS experiment, in which the operator has multiple objectives to perform [24, 64]. Utilization is measured by calculating the ratio of the total service time for all events to the total mission time. In this experiment, utilization accrues anytime the operator is in the SCT window, target identification window, search task window, missile launch approval window, or reading or answering a chat box message. Three types of utilization are explored in this study: (1) required utilization, or the percentage of mission time the operator spends performing mandatory tasks required by the system; (2) self-imposed utilization, or the percentage of mission time the operator spends doing tasks that are the operator's prerogative; (3) total utilization, which is also the sum of required and self-imposed utilization. In addition, a self-rated busyness 5-point Likert metric was collected as a subjective measure of workload.

3.4.2.2 Performance Metrics

The following twelve dependent variables measuring various forms of performance are classified into evaluation categories for human-automation performance metrics [65]. Each dependent variable is organized by human supervisory control metric class and described. The dependent variables for this experiment are well-rounded since all metric categories are represented.

Mission Effectiveness

The mission effectiveness metrics are the three primary performance measures of this experiment because they represent the key mission parameters of search, track, and destroy.

- Target Finding Score: speed of finding targets and quantity of targets found.

Target finding score is calculated as follows:

$$\frac{\sum_{i=1}^{i} \frac{d_i}{a_i}}{F} \tag{1}$$

where

d = time to detect a target

a = time target was available to be found

F = number of targets found

i = a target that was found; $1 \le i \le 4$

This equation yields scores ranging from 0 to 4, where a lower value is better. Four is the worst possible target finding score. The target finding score is computed using this equation when a participant finds between 1 and 4 targets. If the participant finds no targets, that participant receives a score of 4.

- Target Tracking Percentage: percentage of time targets are tracked.

Target tracking percentage is calculated as follows:

$$\frac{\sum_{i=1}^{i} t_i}{\sum_{i=1}^{i} a_i} \tag{2}$$

where

t = total time a target was tracked

a = time target was available to be tracked

i = a target that was found; $1 \le i \le 4$

This equation yields percentages between 0% and 100%, where 100% is the best possible continuous target tracking percentage. If a participant finds no targets, that participant receives a target tracking percentage of 0%.

- Hostile Destruction Score: speed and quantity of hostile destructions.

Hostile destruction score is calculated as follows:

$$\frac{\sum_{i=1}^{i} \frac{d_i}{a_i}}{D} \tag{3}$$

where

d = time to destroy a hostile

a = time hostile was available to be destroyed

D = number of hostiles destroyed

i = a hostile that was destroyed; $1 \leq i \leq 2$

This equation yields scores ranging from 0 to 2, where lower is better. Two is the worst possible hostile destruction score. The hostile destruction score is computed using this equation when a participant finds between 1 and 2 targets. If the participant destroys no hostiles, that participant receives a score of 2.

Human Behavior Efficiency

Each of the following metrics represents information processing efficiency:

- Average Prompted Search Reaction Time: average time to create a search task after chat box prompting
- Average Chat Reaction Time: average time to answer a chat box question
- Average Replan Reaction Time: average time to click on the blinking replan button when prompted by the automation

Human Behavior Precursors

The following cognitive precursors measure situational awareness:

- Chat Accuracy: percentage of correct answers to chat box mission awareness questions
- Prompted Search Accuracy: percentage of correctly placed prompted search tasks

Collaborative Metrics—Human/Automation Collaboration

Each of the following metrics falls into the collaboration with automation category because they represent extra, operator-driven events that involve interaction with the automation. The participants chose to interact with the automation more than required, which indicated desire to collaborate with the automation.

- Number of Search Tasks Created: total operator-created search tasks
- Extra Search Tasks: total operator-generated search tasks; not chat box prompted
- Extra Replans: total operator-generated replans; not prompted by the automation
- Extra Target Edits: total operator-generated uses of the target identification window

3.4.2.3 Attention State Metrics

Video data provided a means of measuring the participants' attention states at all times during the experiment test session. Each participant's time was categorized 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 attention states are further subcategorized into physiological, social, or cognitive. The criteria for video coding the participants' time into these categories 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 in the following ways.

2p). Physiological diversions (examples: yawning, eating, fidgeting, stretching, and scratching)

2s). Social diversions (examples: talking, glancing at each other)

2c). Cognitive diversions (playing Minesweeper or flash games on the same screen as the simulation interface)

3). Distracted Attention

The participant is not paying attention to the interface at all.

3p). Physiological distractions (examples: sleeping, eating a meal without looking at the interface)

3s). Social distractions (examples: discussions with participants' backs turned to the computer)

3c). Cognitive distractions (reading a book, using the internet or other applications on the second screen, checking email and phone messages without looking back at interface)

Video coding software was used to take notes on how each participant allocated his or her attention throughout the 4-hour test session. The instant a participant began performing a particular action, a time-stamped note was taken to categorize the action into one of the aforementioned attention states. The video coding method produced 100% agreement across 3 raters for 5/30 video files due to the objective, rule-based rubric. The time between time stamps was counted as the amount of time the participant was in that particular attention state.

3.5 Methodology Summary

The OPS-USERS testbed was altered to create a long duration, low taskload scenario. Experimental data was collected for 30 participants of ages comparable to military unmanned vehicle operators, including metrics of workload, performance, video data, and demographic data, which included a self-assessment of gaming experience and comfort level with computer programs. Three participants performed their supervisory control missions at the same time in a simulated control room that had possible distractions, including each other. The independent variable for controlling the experiment was the replan interval, which was the time participants were prompted to evaluate a plan generated by the automation. The primary performance metrics focused on search, track, and destroy speed and quantity. Other performance metrics included reaction times and accuracies to prompted events. Extra instances of interacting with the automation were also measured to gauge self-imposed types of workload. The results of this experiment will be discussed in the next chapter.

4 Results and Discussion

This chapter discusses the impact of the long duration, low workload experiment on operators' utilization, attention, and performance. The statistical results from the analysis are provided, followed by discussion. This chapter addresses the five research hypotheses: (1) the performance of operators at low workload will not follow the Yerkes-Dodson curve; (2) taskload will have a positive, linear relationship with utilization; (3) boredom will affect attention state by decreasing directed attention; (4) directed attention will improve performance; and (5) vigilance will decrease over time. Each of these five hypotheses corresponds to the five main research questions. In addition, a top performer analysis is discussed. Finally, a performance comparison is made between this low workload experiment and the previously-discussed moderate workload study. Overall, this study seeks to determine how human subjects behave under long duration boredom in a multi-UxV mission.

4.1 Utilization

The first two research questions investigated in this study involve utilization, or the percent busy time, excluding monitoring time. The Yerkes-Dodson curve predicts that performance degrades as workload decreases [20]. The first research question sought to determine whether the Yerkes-Dodson curve prediction is accurate, specifically in human supervisory control situations of low workload. It was hypothesized that the performance curve will become horizontal as the curve approaches the lowest workload. The second related research question considered how taskload affects operator utilization, the workload metric. It was hypothesized that taskload would affect utilization with more taskload causing higher utilization.

To test both of these utilization hypotheses, the experimental control for workload involved 3 levels of required utilization, modulated by the independent variable, replan interval. Participants replanning at the 10-minute replan interval were required to replan twice as often as the 20-minute interval group and three times more frequently than the 30-minute interval group. 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%.

Even though participants were grouped into 3 different levels of workload, an interesting result occurred; regardless of the fact that some participants were given more taskload than others, they all gravitated to the same narrow range of utilization: an average of 11.4% with a standard deviation of 3.36%. A non-parametric test, the Kruskal-Wallis test, showed that utilization was not statistically different across the 3 replan intervals ($\chi^2 = 0.135$, $p = 0.935$). Hence, utilization was not dependent on replan interval. Due to the extremely low workload nature of the study, participants interacted with the simulation as much as they pleased, regardless of the lower required utilization controlled by certain replan intervals.

Since the replan interval groups did not have significantly different utilizations, the low workload end of the Yerkes-Dodson curve was neither confirmed nor disconfirmed by the experimental design for this research. Figure 12 shows the average utilization and overall performance for all 30 participants. The overall performance

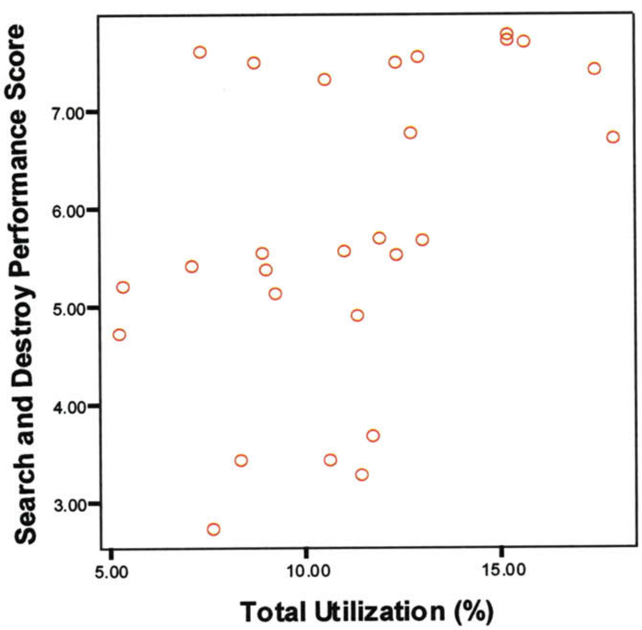

Figure 12: Utilization versus Performance

metric is based on target finding score (Equation 1) summed with hostile destruction score (Equation 3) and normalized so that a higher performance value is better with 8 being the highest possible score. The search and destroy performance metrics were chosen to represent participant performance because the system performance in these tasks depends most on operator interactions. Target tracking is highly automated and is not included in measuring human performance. The data in Figure 12 did not confirm the inverted "u-shape" curve for utilization versus performance as the Yerkes-Dodson curve suggests, due to the large variability in performance scores.

A deeper investigation of utilization was necessary to determine why operators gravitated to a common utilization in the long duration, low taskload environment. All participants purposely over-utilized themselves by interacting with the system more than the mission requirements dictated. This over-utilization may be due to the extra cognitive capacity that the participants had during the low workload scenario. The important aspect of this finding is that utilization can be categorized into two subcategories of utilization: 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 that were prompted by the command center, number of targets found that required identification, and number of hostiles destroyed that required operator approval. Each participant's required utilization was specific to the replan interval independent variable. However, even participants who were required to replan at the same intervals had different required utilizations because each participant had a slightly different situation based on how many targets they found, how many hostiles they destroyed, and how long they spent performing each event.

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.

Self-imposed utilization activities include extra replanning, creating participant-generated search tasks, and additional uses of the target identification window for editing target designation.

On average, participants were required to be 1.87% utilized (s.d. 0.49%), yet the average total utilization was 11.4% (s.d. 3.36%). The average self-imposed utilization was 9.53% (s.d. 3.33%), which is five times more utilization than required by the mission scenario.

As with total utilization, participants gravitated toward the same level of self-imposed utilization. The Kruskal-Wallis test showed that self-imposed utilization was not statistically different across the three replan intervals ($\chi^2 = 0.439$, $p = 0.803$). However, the three different replan intervals caused significantly different required utilization ($\chi^2 = 16.579$, $p < 0.001$). The 10-minute interval group had an average of 2.41% required utilization (s.d. 0.46%), the 20-minute interval group had an average of 1.69% required utilization (s.d. 0.14%), and the 30-minute interval group had an average of 1.58% required utilization (s.d. 0.36%). The bar chart in Figure 13 shows the average amount of total utilization, categorized into self-imposed and required utilization, for each of the three replan intervals.

The 10-minute interval group had the highest required taskload and the 30-minute interval group had the lowest required taskload. In effect, the independent variable caused different levels of required utilization, but not total utilization. The hypothesis that taskload will affect utilization only holds true for required utilization ($\chi^2 = 16.579$, $p < 0.001$), but not for total utilization ($\chi^2 = 0.135$, $p = 0.935$) or self-imposed utilization ($\chi^2 = 0.439$, $p = 0.803$).. Rather, because participants engaged in self-imposed utilization, the total utilization was not affected by replan interval. At low taskload,

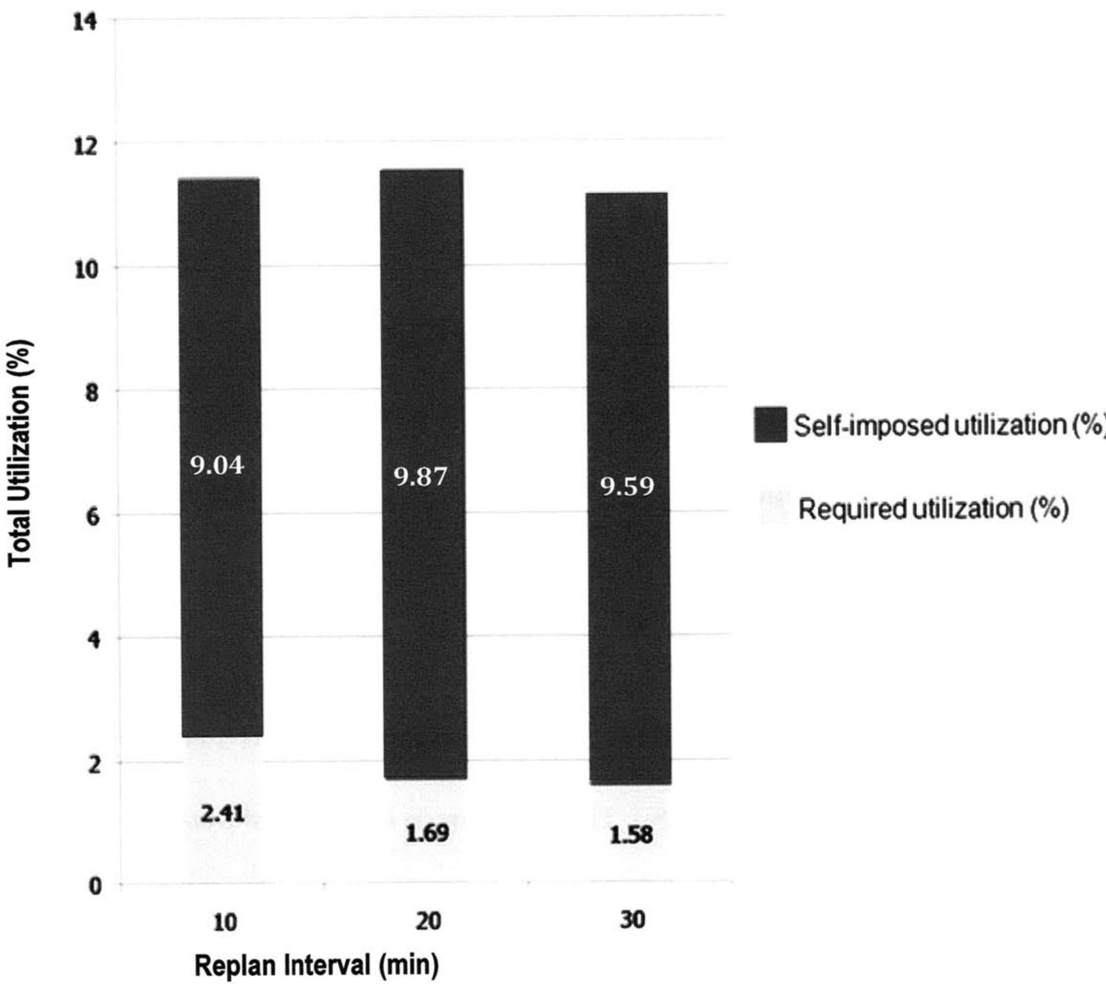

Figure 13: Utilization by Replan Interval

operators created extra work for themselves. This finding shows that humans do not comfortably operate at low workload and that they crave at least a moderate level of workload to sustain their attention.

4.2 Attention

The third research question investigated how the low workload environment affected operator attention. The associated hypothesis predicted that operators would spend most of their time in divided attention (in an effort to continue paying attention but coping with the boredom by multitasking), some of their time completely distracted (due to boredom), and the least amount of their time in directed attention (because of low workload and interest).

As described in Chapter 3, directed attention is the amount of time that participants directed their attention toward the interface. Divided attention represents time that the participants spent multitasking physically (such as eating or stretching), socially (such as talking over their shoulder or quickly glancing at one another), or cognitively (such as playing Minesweeper on top of the interface). All divided attention state subcategories involve participants maintaining visual contact with the interface and paying attention to the mission in some capacity. Anytime the participants were not looking at the interface is considered distracted in one of three categories: physically (such as sleeping or going for a snack), socially (such as talking to each other or on the phone with their backs toward their interfaces), or cognitively (such as reading, texting, playing games, checking email, or browsing the internet). All of these coping actions occurred at least once.

Video coding analysis showed 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. Figure 14 illustrates the average attention allocation of participants during the long duration, low workload experiment.

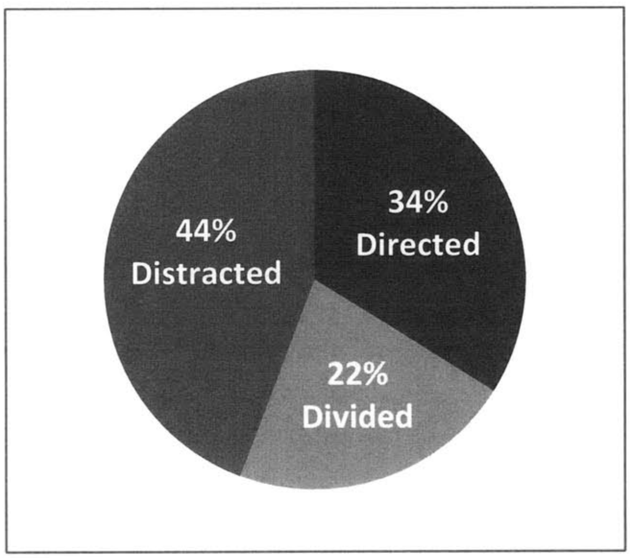

Figure 14: Attention State Allocations

The non-parametric Friedman test showed that these three percentages of attention allocation are statistically different (χ^2=8.267, p = 0.016). Three more tests were run on this attention allocation data to determine the pairwise comparisons, making the family-wise error value α = 0.026 for significance. The aforementioned Friedman test met this threshold for significance. The Wilcoxon Signed Ranks test was used compare the attention states. The results are shown in Table 1.

Table 1: Attention State Pairwise Comparisons

Attention State Comparison	Z	p
Directed > Divided	-2.828	0.005
Distracted > Divided	-3.260	0.001
Distracted > Directed	-1.656	0.098

The pairwise comparisons involving divided attention are clearly statistically significant because they not only meet the $\alpha = 0.1$ for non-parametric testing but also the family-wise error $\alpha = 0.026$. On the other hand, the comparison between distracted and directed attention only meets the $\alpha = 0.1$ significance level for non-parametric testing. Overall, it is seen that participants spent significantly different amounts of time among the three primary attention states.

These attention state allocation results did not match the hypothesis that participants' attention would be allocated in order from highest to lowest: divided, distracted, and then directed. In fact, directed attention was not the lowest amount of attention; divided attention was the least likely, and participants spent the least amount of time multitasking. While enduring such a long duration, low workload simulation, it is surprising that participants were able to spend so much of their time in directed attention toward the simulation. The $250 Best Buy gift card reward enticed the participants to put forth more effort than expected in this boredom study. However, participants were distracted for the majority of the time, and divided attention in multitasking was the least likely attention state.

The descriptive statistics of the sub-categories of the 3 attention states are shown in Table 2.

Table 2: Attention State Descriptive Statistics

Attention State	N	Minimum	Maximum	Mean	Std. Deviation
Divided Socially	30	.00	.10	.03	.03
Divided Physically	30	.03	.55	.17	.13
Divided Cognitively	29	.00	.13	.01	.03
Distracted Socially	30	.00	.29	.09	.09
Distracted Physically	30	.00	.18	.06	.05
Distracted Cognitively	30	.04	.59	.29	.15
Total Directed	30	.10	.75	.34	.15
Total Divided	30	.09	.55	.22	.13
Total Distracted	30	.07	.79	.44	.20
Valid N (listwise)	29				

Overall, participants spent more time in a distracted state than any other attention state. The mode distraction subcategory was cognitively distracted with a mean of 29% (s.d. 15%). Participants were much more likely to be using their cell phones, doing homework, checking their email, or reading a book than talking to each other, eating, or sleeping. Second to distracted attention was purely directed attention with a mean of 32% (s.d. 15%). Below directed attention, the subcategory of "divided physically" was most prevalent, with a mean of 17% (s.d. 13%). When multitasking, participants stretched, shifted in their seats, and snacked much more than talking or playing a cognitive game while still looking at the interface. Examining how participants allocated their attention tells a great deal about how a long duration, low workload mission affects the human operator.

Performance can be predicted from attention allocation. The fourth research question asked whether performance can be predicted by attention states in a low workload environment. The hypothesis was correct in that participants with more directed attention performed better. The scatter plot in Figure 15 illustrates the positive trend between directed attention and performance in search and destroy tasks.

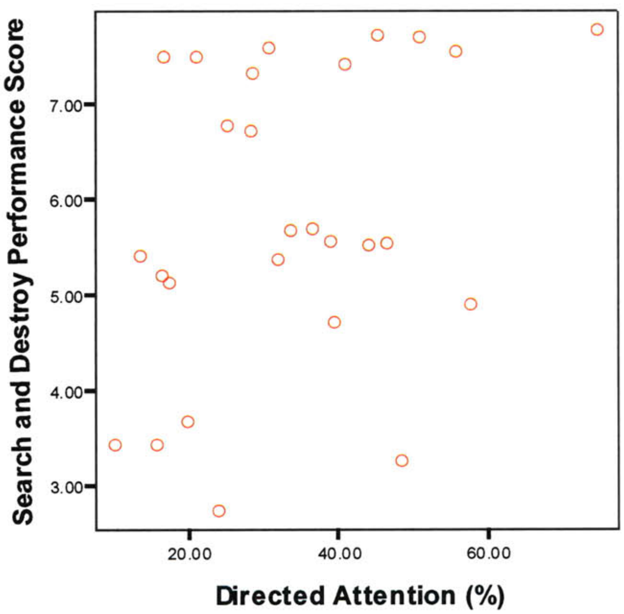

Figure 15: Directed Attention versus Performance

There is a marginally significant correlation between directed attention and performance (Spearman's $\rho = 0.372$, $p = 0.056$). This finding is important because it shows that performance in long duration, low workload environments can be improved with higher levels of directed attention. In addition, directed attention is highly correlated with total utilization (Pearson's $\rho = 0.434$, $p = 0.017$), as shown in Figure 16. Thus, in a low taskload environment, more utilization, or workload, may be the key to more directed attention, and hence, better performance.

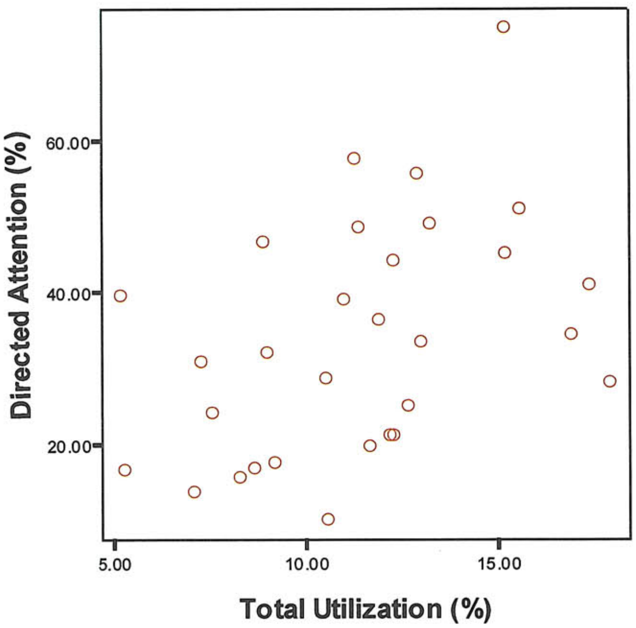

Figure 16: Utilization versus Directed Attention

4.3 Performance

The fourth hypothesis, discussed in the previous section, supposed that performance could be predicted in the low workload environment. To further investigate this performance prediction, three linear regressions were calculated, one for each of the 3 primary performance metric categories: search, track, and destroy. These linear regression models were generated using the backward selection method. The predictor variables include total utilization, percent directed attention, and gaming level. The model coefficients and significance levels are shown in Table 3 and will be discussed in the following sections for each primary performance metric.

Table 3: Linear Regressions

Performance Metric	R^2	β_0	Total Utilization	Directed Attention	Gaming Level
Target Finding Score	.254	$\beta = 0.906$ $p < .001$	$\beta = -4.282$ $p = .007$	N/A	N/A
Target Tracking Percentage	.189	$\beta = 0.998$ $p < .001$	$\beta = -.637$ $p = .048$	$\beta = 0.131$ $p = .049$	N/A
Hostile Destruction Score	.326	$\beta = 1.177$ $p = .032$	$\beta = -9.055$ $p = .015$	N/A	$\beta = 0.518$ $p = .038$

The corresponding Kolmogorov Smirnov tests for normality and Levene tests for homoscedasticity are detailed in Appendix G.

4.3.1 Search Performance Prediction

The target finding score metric incorporates the speed and quantity of targets found, as detailed in Chapter 3. A lower target finding score indicates better performance.

The linear regression model for target finding score suggested that total utilization is the only predictor variable that influences a person's target finding ability. The model for target finding is represented mathematically in Equation 4,

$$y = 0.906 - 4.282u \tag{4}$$

where y represents target finding score and u depicts total utilization ($p = 0.007$). This model shows that a 1% increase in total utilization lowers the target finding score by

0.04, thereby improving target finding since a lower score is better. This result suggests that more interaction with the simulation in a low workload scenario improves search performance.

Target finding score correlated with hostile destruction score ($\rho = 0.593$, $p = 0.001$). Participants who found many targets and found them quickly also destroyed many hostiles quickly. These search and destroy metrics go hand-in-hand and are more dependent on the human operator than the automation. Targets must be identified by a human operator just as weapons approval must be made by a human operator. On the other hand, target tracking does not necessarily require human interaction with the system to be accomplished. The auto-planner schedules the UxVs to track targets automatically, while the human operator can monitor and approve these schedules. However, the act of tracking a target is not a discrete event in which the human operator participates. The next section on target tracking illustrates how more participant interaction hinders target tracking and simultaneously augments target finding and hostile destruction.

4.3.2 Track Performance Prediction

The target tracking percentage metric is calculated by dividing the total amount of time a participant's UxVs track the emergent targets by the total amount of time the targets were available to be tracked. Before a target has been discovered, it cannot be tracked. The amount of time from target finding to the end of the simulation therefore represents the total time a target was available to be tracked. Target tracking was done automatically by the UxVs. Once a target was designated by the operator as unknown or hostile, the auto-planner put the target into the queue to be tracked automatically. Target tracking is primarily left up to the automation after the operator identifies an emergent target and accepts a schedule that assigns that target in the SCT. Target tracking involves revisiting the moving target often enough that the target does not become "lost." A lost target is one that is not found again at its last known location nor at its projected location based on the targets last known velocity vector and time since target sighting. The average number of targets participants lost was 0.93 (s.d. 1.2 targets).

The linear regression for target tracking percentage showed that a participant's total utilization and percentage of directed attention both predict the system's ability to track targets, as shown in Table 3. The model for target tracking is

$$y = 0.998 - 0.637u + 0.131d \tag{5}$$

where y represents target tracking percentage, u stands for total utilization, and d depicts the directed attention state. The first significant predictor of target tracking is total utilization with p = 0.048. A 1% increase in total utilization results in a 0.637 percent decrease in target tracking. The more a participant interacted with the simulation, the worse the target tracking became since the automation is not left alone to operate optimally in target tracking. This interruption caused a lag in automated target tracking assignments to the UxVs, decreasing the target tracking percentage.

The second predictor of target tracking is percent directed attention with $p = 0.049$. A 1% increase in percent directed attention causes a 0.131 increase in percent target tracking. Even though tracking is considered primarily automation-driven, having an operator intently monitor the system to make sure targets are not becoming lost ameliorates target tracking.

The extra target edits variable was not included in the linear regression because it correlates with the predictor variable total utilization ($\rho = 0.392$, $p = 0.035$). The correlations of extra target edits show that participants who over-interacted with the system by editing targets beyond the system requirements had worse target tracking.

Target tracking works best when the automation is left alone, yet monitored by a human supervisor. Although target tracking is automated, directed attention nonetheless assists the system in not losing targets because a human operator can intervene with search tasks according to the situation. However, directed attention improves target tracking percentage less than $^1/_5$ as much as a lack of utilization does. However, the overall mission would be impossible without the necessary operator interactions for destroying hostile targets, as discussed in the next section. A balance must be struck for overall mission performance; although operator interaction via utilization hinders target tracking, it advances both the search and destroy tasks.

4.3.3 Destroy Performance Prediction

The hostile target destruction metric is calculated similarly to the target finding score. Hostile destruction score incorporates the speed and quantity of hostiles destroyed. A lower hostile destruction score indicates better performance.

The linear regression for hostile target destruction is predicted by total utilization and gaming level. The model is:

$$y = 1.177 - 9.055u + 0.518g \qquad (6)$$

where y is the hostile destruction score, u represents total utilization, and g signifies gaming level. The first significant predictor variable for hostile destruction is total utilization ($p = 0.015$), just as for target finding score. A 1% increase in total utilization results in a 0.09055 decrease in hostile destruction score, which is an improvement. The more interaction participants have with the simulation, the faster all the hostiles can be destroyed. Thus, keeping the human interacting with the system is key to good performance in hostile destruction.

The second predictor variable for hostile destruction score is gaming level ($p = 0.038$). An increase in experience level from non-gamer to gamer results in a 0.518 increase in hostile destruction score, which is a large decrement in hostile destruction performance. This finding suggests that gamers are not well-suited for long duration, low workload missions in supervisory control because of their conditioned need for stimulus. The task of approving weapons launches mimics the exciting missions of video games; however, when combined with a low workload environment, the task of approving weapons launch does not bring out the best performance in gamers.

Extra replanning events also correlated with improved hostile destruction (ρ = -0.432, p = 0.025). Extra replans involve more interaction with the system, or total utilization, and increase hostile destruction performance. Extra replanning was not included in the linear regression because it correlates with total utilization (ρ = 0.577, p = 0.001). In addition, hostile destruction score correlated strongly with target finding score (ρ = 0.593, p = 0.001). Participants who found many targets quickly also had a tendency to destroy many hostiles quickly.

In terms of information processing and situational awareness, hostile destruction performance negatively influenced attending to automation-prompted search tasks. Hostile destruction score correlated negatively with increased prompted search task average reaction time (ρ = -0.396, p = 0.046). In addition, hostile destruction performance correlated with poorer prompted search task accuracy (ρ = 0.408, p = 0.035). Participants were so focused on destroying a hostile target that they neglected their duties of quickly and accurately creating search tasks when prompted.

4.4 Attentional Effects on Operator Behavior

Correlations among performance metrics other than search, track, and destroy tasks present some interesting research findings. First, attention state affects utilization, and therefore performance. Total directed attention correlated with extra search tasks (ρ

= 0.509, p = 0.004) and extra replans ($\rho = 0.580$, $p = 0.001$) just as total divided attention correlated with extra search tasks ($\rho = 0.453$, $p = 0.012$) and extra replans ($\rho = 0.374$, $p = 0.042$). Oppositely, total distraction correlated negatively with extra search tasks ($\rho = -0.684$, $p < 0.001$) and extra replans ($\rho = -0.689$, $p < 0.001$), since a participant cannot interact with the interface when they are not looking at it. These correlations make it clear that attention state does, in fact, affect behaviors that add to utilization.

The directed and distracted attention states correlated with utilization that influenced performance. Total utilization correlated with total directed attention ($\rho = 0.434$, $p = 0.017$). Self-imposed utilization correlated negatively with total distraction ($\rho = -0.406$, $p = 0.026$). The more utilization a participant self-imposed, the less likely they were to be completely distracted. One way for participants to have less distracted attention and possibly more directed attention was to engage in self-imposed utilization. More directed attention led to higher utilization and better performance, whereas self-imposed utilization prevented distraction.

This long duration, low workload study showed that performance in creating search tasks and chat messaging suffered, even with increasing utilization. As discussed previously, increasing total utilization improved performance in the primary mission tasks of search and destroy. Interestingly, chat response accuracy negatively correlated with total utilization ($\rho = -0.498$, $p = 0.005$). The more a participant interacted with the system, the less accurate their responses were to the command center situational awareness questions. It is surprising that a low workload study with such a low average total utilization (11.4%, s.d. = 0.03) could show a decrease in situational awareness as utilization increases. In addition, even at a low workload setting, participants' reaction times slowed with increasing levels of required utilization. Required utilization correlated with prompted search average reaction time ($\rho = 0.439$, $p = 0.015$) and chat average reaction time ($\rho = 0.502$, $p = 0.006$), which suggests that the more the required utilization increased, the worse the reaction times became. Conversely, in the previously discussed moderate workload study, increasing utilization did not significantly correlate with worsened reaction times. The poor performance in reaction times only occurred in the low workload study. Malleable attentional resource theory explains that performance often suffers in situations of mental underload [66], and the lengthened reaction times and worsened chat response accuracies of this low workload experiment illustrate this point. Ordinarily, a decrease in task performance constitutes a limit in mental capacity. However, the low taskload imparted on participants and the low levels of utilization measured show that they were clearly not overloaded, but perhaps the boredom did cause their mental capacity to be filled.

Other correlations demonstrated that participant behaviors in different tasks could cause a snowball effect of good performance. Prompted search task average reaction time and accuracy, while different metrics of different categories (e.g. information processing and situational awareness), were strongly correlated ($\rho = -0.801$,

$p < 0.001$); this is a positive correlation because a lower reaction time is better. Participants created prompted search tasks with equal measures of speed and accuracy. If participants attended to the task quickly, they were also likely to be accurate. Likewise, participants who made copious amounts of extra search tasks were also likely to engage in many extra replans, as shown in the correlation between extra replans and extra search tasks created ($\rho = 0.914$, $p < 0.001$). Extra search tasks and replans all increased total utilization, which was shown to improve performance.

4.5 Vigilance Degradation

The final research question considered whether vigilance degrades over time in a long duration, low workload mission involving human supervisory control of networked UxVs. It was hypothesized that directed attention would degrade over time. This hypothesis was supported. A Repeated Measures General Linear Model showed a significant difference in directed attention across hour intervals ($F = 21.953$, $p < 0.001$). Tukey pairwise comparisons showed a statistical difference at the $\alpha < 0.05$ level in directed attention between all hour intervals, except the comparison between the third and fourth hour. The second hour was also only marginally different from the fourth hour ($p = 0.066$). The p values for all comparisons can be referenced in Appendix H. Figure 17 shows the estimated means plot of how vigilance decreases overtime. The error bars show standard error. Note that hours 3 and 4 are not statistically different, even though the amount of directed attention appears higher in hour 4. Directed attention starts out high and decreases, eventually flatlining from hours 3 to 4.

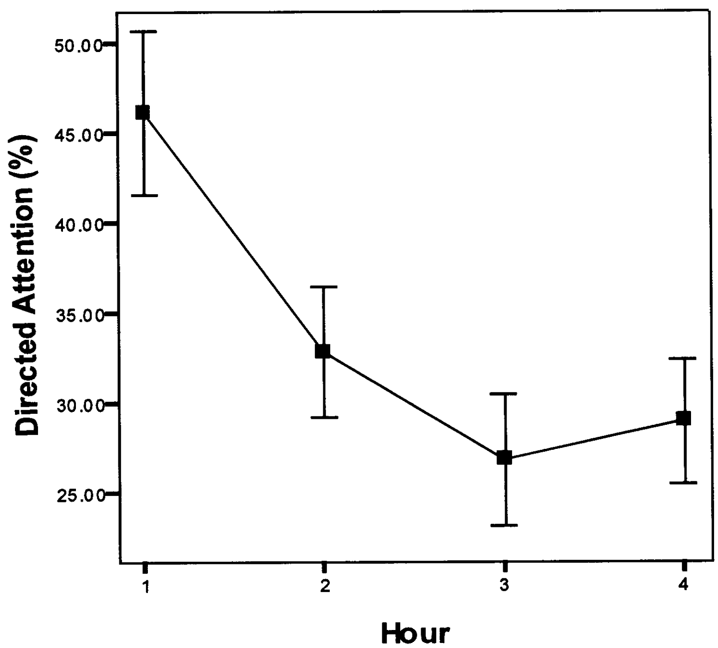

Figure 17: Estimated Means Plot for Vigilance Degradation

4.6 Research Question Summary

This research showed that performance does not necessarily decrease with low workload, especially in the context of human supervisory control of networked UxVs. Given varying levels of low taskload, operators tended to gravitate toward a common total utilization that was well above the required utilization. The boredom caused by the low workload environment caused operators to spend the majority of their time in distracted attention, followed by directed attention, and the least amount of time multitasking in divided attention. More directed attention predicts higher operator performance, especially in the tracking portion of the mission.

Higher utilization predicts improved operator performance in search and destroy tasks, but hinders the automation's ability to track targets. Gaming experience was a detriment to destroying hostile targets in this long duration, low workload mission. Vigilance, shown by a decrement in amount of directed attention per hour, decreased over the course of the mission duration. The descriptive statistics for all data gathered can be found in Appendix I. Sources of error are listed in Appendix J. The next section describes the coping mechanisms of the top performers.

4.7 Top Performer Analysis

This section describes the top 8 performers and gives insight into how participants coped with the low workload in order to outperform the majority of participants. The top 8 performers were identified as having a standard deviation of at least 1 below the mean performance score, where a lower performance score is better. Figure 18 shows the mean as a solid line and one standard deviation below the mean as a dashed line.

Performance score is based on the target finding score and hostile destruction score, which were detailed in Chapter 3. Although the mission involves all three categories of search, track, and destroy, only search and destroy truly measure human performance, whereas the track metric is a better measure of automation

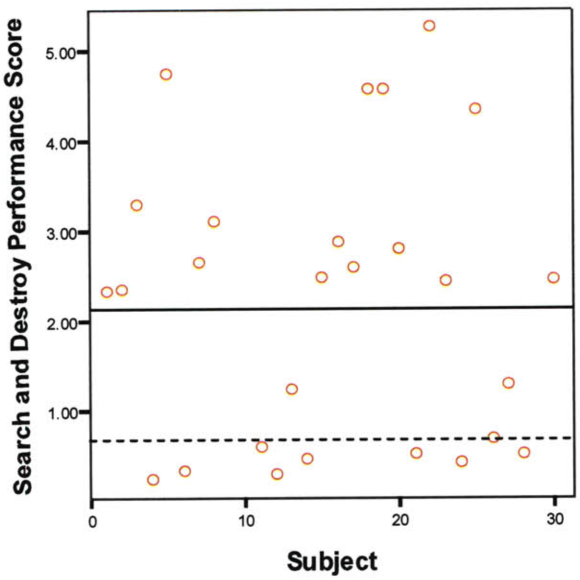

Figure 18: Top Performer Selection

performance. Thus, the target finding and hostile destruction score were represented in the total performance score. Since target finding score is on a scale from 0 to 4 and the hostile destruction score is on a scale from 0 to 2, the hostile destruction score was doubled to be on an equivalent scale as the target finding score. These two scores were summed to obtain the performance score where lower is better. The top 8 performers' scores ranged from a high score of 0.23 and a low score of 0.59.

These participants were analyzed to further understand how humans can succeed in a long duration, low workload mission. Six of the 8 top performers were non-gamers, whereas only 2/8 were gamers. It is interesting that the 2 gamers of the top performers were both female. Six of the 8 top performers had military experience, and only 2/8 were not in the military. It is interesting that so many top performers were in the military since only 43% of participants overall had military experience.

The top performers included 4 males and 4 females. Thirty-six percent of all females who participated in this long duration, low workload experiment were top performers. Only 21% of males who participated in this experiment were identified as top performers. Future research should be conducted to validate whether women are better at sustained alertness tasks.

The winner of the experiment was a 19-year-old female college student with no military experience who does not play video games. It can be immediately deduced that this description of the top supervisory controller of networked UxVs does not match current stereotypes of the military's UxV pilots for search, track, and destroy missions. The winner, the youngest participant, had a total utilization of 15.2%, although she was only required to be 1.6% utilized. In the post-experiment survey, she reported feeling busy, self-rating a 3 out of 5 busyness level. Of all the top performers, the winner felt the busiest. It is interesting that the winner had a neutral perception toward UxVs and also indicated a low comfort level with using computer programs. Her conscientiousness helped her. She had a middle-of-the-road self-rated confidence score of 3 out of 5, although most of the top performers felt very confident with a median self-rated confidence of 4 out of 5. The winner's self-rated performance was "good," or 4 out of 5, like most of the top performers. One of the top performers did indicate a self-rating of excellent performance (5 out of 5). Appendix K shows the demographic and post-experiment survey data for the top performers. Figure 19 shows a bar graph of top performers' self-rated confidence and self-rated performance with the performers listed in order of performance.

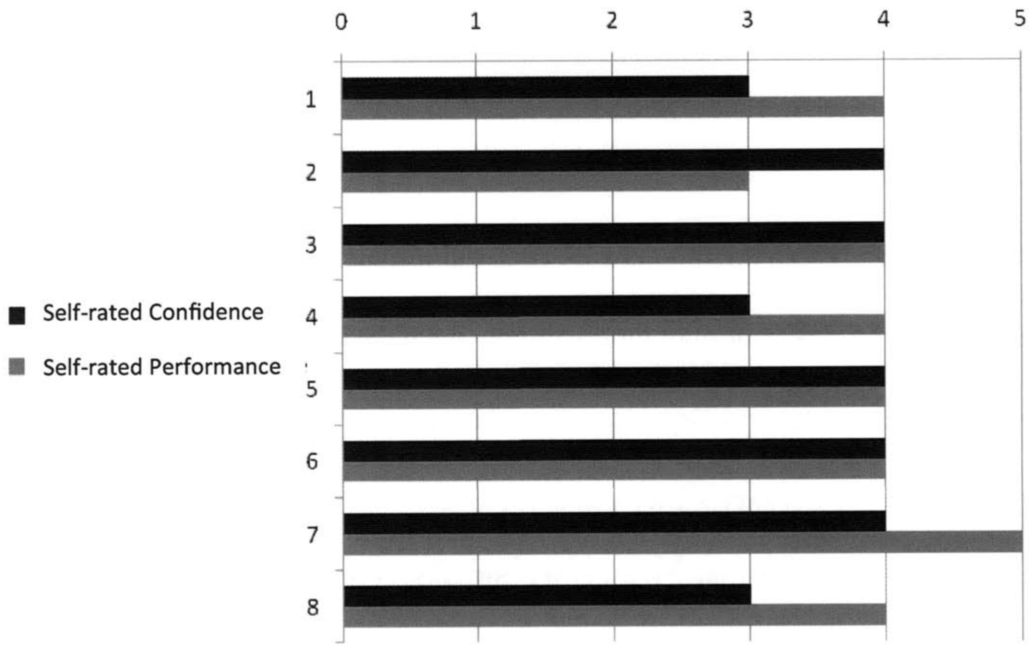

Figure 19: Confidence and Performance Self-Ratings

The characteristic of the winner that set her apart was her extremely high amount of directed attention; she appeared focused 75% of the time, whereas the average amount of directed attention for all the top performers was 41% (s.d. = 20%), and the overall average of directed attention was only 34% (s.d. = 15%). Thus, the top performers' average directed attention was higher than the overall average 34% (s.d. = 15%). However, 3 of the top performers had below average directed attention, at 31%, 21%, and 17%, yet still managed to be ranked as top performers. The attention state descriptive statistics are shown in Table 4, listed in decimal form. The attention state values represent the percentage of time the participants spent in each state.

Table 4: Attention State Descriptive Statistics for Top Performers

	N	Minimum	Maximum	Mean	Std. Deviation
Divided Socially	8	.00	.07	.02	.02
Divided Physically	8	.03	.25	.14	.08
Divided Cognitively	8	.00	.06	.01	.02
Distracted Socially	8	.01	.19	.05	.06
Distracted Physically	8	.03	.18	.07	.05
Distracted Cognitively	8	.04	.59	.29	.21
Total Directed	8	.17	.75	.42	.19
Total Divided	8	.09	.26	.17	.07
Total Distracted	8	.12	.70	.41	.20
Valid N (listwise)	8				

The top performers operated in different types of social environments. For instance, the winner of the experiment was in a test room that was completely silent because her group members were seemingly introverted like herself. She hardly spoke a word and remained almost entirely focused on the mission simulation. One of her group members fell asleep for nearly half an hour, and neither of her group members were top performers.

A different example shows two of the top performers were in the same test session together, a session in which an intense political debate was going on for a large portion of the mission duration, approximately 120 minutes. One participant became a top performer by ignoring the two group members engaged in the political debate and quietly focusing on the mission (with 41% of her time in directed attention) or by keeping herself alert by reading a book (with 36% of her time in divided attention). The other top performer from that same test session engaged in the political debate the whole time and spent nearly 40% of the time distracted from the mission by talking with the third group member with his back to the computer interface. However, this participant performed extremely well in spite of the high distraction level, and in fact, he was the second place performer of the study. He was able to accomplish excellent performance despite his high distraction in the political debate since he still spent 45% of the time in directed attention, attending to his simulation at frequent intervals during the debate. On average, he attended to his mission 42 times per hour during the political debate, or approximately 84 times during the two-hour debate. The effects of these switching times, going back and forth between the low workload mission and intense debate, was an effective strategy for him in dealing with boredom.

The third group member, who was the instigator of the social debate, was not a top performer because she did not attend to her mission much at all while talking. Whenever the other debater would switch from their discussion to attend to his interface, she would also look away as is the social pattern when someone a person is conversing with directs his attention elsewhere. However, instead of attending to her own mission, the third group member looked at a project on her personal laptop. In essence, the third group member had two sources of distraction, whereas her debate partner only switched between the debate and his mission.

All in all, about half of the top performers were in social environments where the participants conversed throughout the mission, and the other half operated in rooms that had a quiet atmosphere of silent tension. It did not matter which type of environment a participant ended up fostering or being subjected to; a participant could be a top performer whether by talking or being quiet, depending on how they attended to their mission. Either the talking or the silence could have been a coping mechanism.

Participants may have been using two different types of attentional mechanisms to cope with their boredom environment: endogenous and exogenous attention. Endogenous attention involves actively self-sustaining attention on a task one considers important. This typically top-down controlled mechanism requires attentional [49] effort. On the other hand, exogenous attention is an automatic attraction of attention that comes from an outside stimulus or change in stimulus. Exogenous attention [49] functions in a bottom-up manner and is not under a person's voluntary control. Both of these attentional orientations [49] were manifested in this study and helped participants perform the mission. People's different personality types and attentional dispositions may have influenced the way in which they allocated their attention to complete the experiment mission. Personality characteristics could be a facet of future work for understanding how human supervisory controllers cope with low workload. Table 5 provides information concerning the top 8 performer's characteristics, where the category "Social" indicates whether the test group was one that had social interaction as opposed to silence.

Table 5: Top Performer Characteristics

Rank	Score	Directed	Divided	Distracted	Utilization	Female	Military	Social	Gamer
1	0.23	75%	13%	12%	15%	✓			
2	0.28	45%	18%	37%	15%		✓	✓	
3	0.31	51%	26%	23%	16%	✓	✓		
4	0.41	31%	26%	43%	7%		✓		
5	0.46	56%	9%	35%	13%		✓	✓	
6	0.51	21%	9%	70%	12%	✓	✓		✓
7	0.51	17%	15%	69%	9%		✓	✓	✓
8	0.59	41%	23%	36%	17%	✓		✓	

The defining factor for top performers was either showing exemplary discipline to focus on the mission or else employing strategic switching times between distractions and the mission. Three top performers had below average directed attention and still came out on top because of effective switching times, like the second place winner. It is interesting that this second place winner scored so closely to the first place winner, only differing by 0.05 out of an 8.0 performance scale with 0.0 being the best. The second place winner was the opposite type of person as the first place winner in that he was one of the oldest participants at age 28, male, with military experience, although not a gamer. Instead of using extreme focus to complete the mission, he used switching times between distractions and the mission. It is also interesting that the third place winner scored even closer to the second place winner, only differing by 0.03 out of an 8.0

performance scale. The third place winner was similar to the first place winner in terms of a highly focused strategy. The third place winner also reported feeling busy during the low workload mission. The first and third place participants were both females and the only two to report feeling "busy," while all other participants reported "not busy" or "idle." These first and third place winners outperformed the rest of the participants even with a higher perceived workload.

Overall, this analysis suggests that participants were able to be top performers even though they were distracted on average 43% of the time. In other words, distraction is not necessarily detrimental to mission performance. This research suggests that participants with very high levels of focused attention showed exemplary performance; in addition, participants with moderately high distraction also performed well because of employing effective switching times.

4.8 Performance Comparison with a Moderate Workload Study

In order to determine how well participants in the long duration, low workload experiment performed relative to other multi-UxV studies, a comparison was made between this experiment and the previous replan interval experiment discussed in Chapter 2. The previous experiment researched moderate levels of workload, ranging from 30% to 70% utilization, whereas the utilizations in this experiment ranged from 5% to 18%. The 31 data points for the moderate workload experiment were taken from the 45-second replan interval dataset, given this was the best performance condition, and all 30 data points were used from the low workload study. The performance comparison was made in terms of target finding score and hostile destruction score, the two primary human performance metrics detailed in Chapter 3. These metrics take into account speed and quantity of targets found and hostiles destroyed. In order to compare the two studies, the scores in both of these categories were normalized to the same scale with scores ranging between 0 and 1, where 1 is the best possible score. The target finding score comparison shows that under low workload participants are able to achieve the highest target finding scores as well as the lowest target finding scores. Figure 20 shows these results. The data for target finding appear similar for both studies.

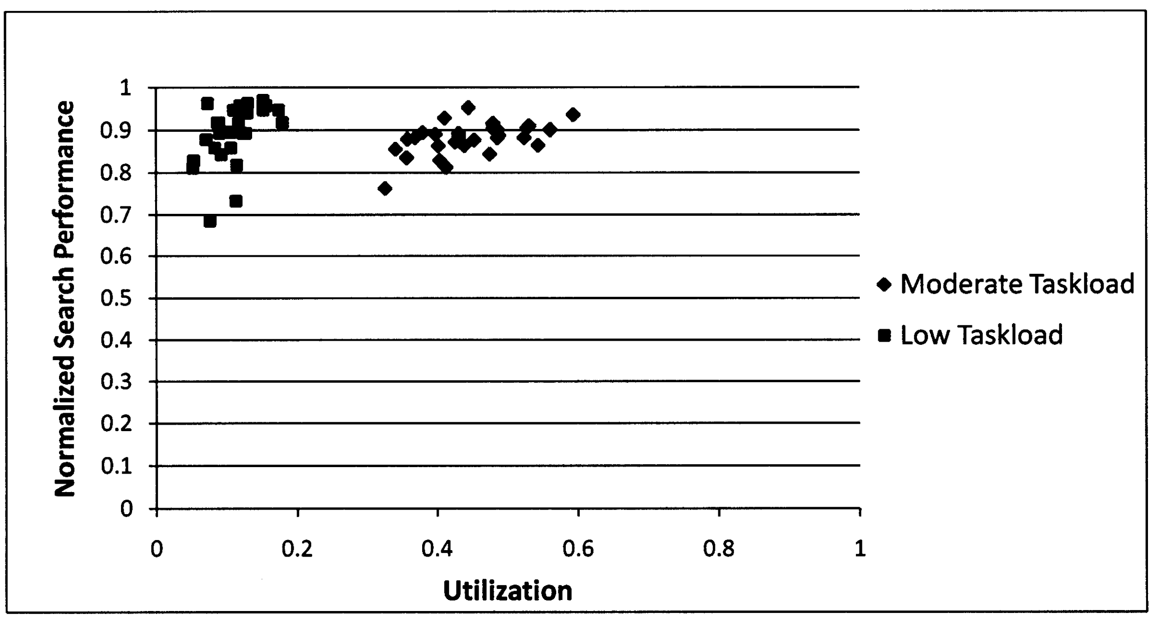

Figure 20: Low Workload versus Moderate Workload in Target Finding

The hostile destruction score comparison shows the same trend; low workload brings both the highest and lowest performance scores, but with more variance in the data. Figure 21 shows the comparison for hostile destruction.

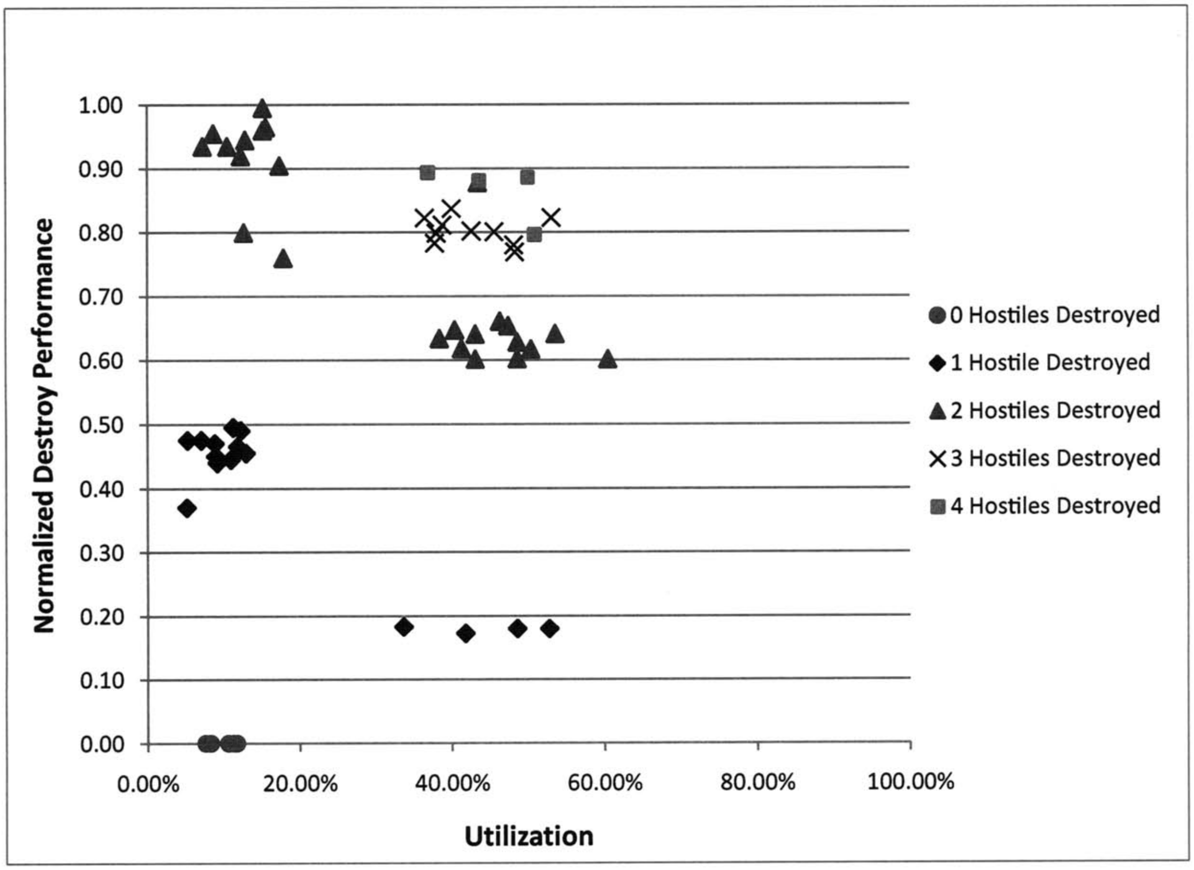

Figure 21: Low Workload versus Moderate Workload in Hostile Destruction

The data appears clustered at discrete levels of performance. This clustering is primarily due to dividing the speed ratio by the integer number of hostiles destroyed. Participants fell into three distinct groups of high, medium, and low performance. Table 6 shows the mean attention allocation of each group. There is a trend of increasing performance with increasing directed attention and decreasing distracted attention.

Table 6: Attention Allocation of Hostile Destruction Groups

Hostile Destruction	Directed	Divided	Distracted
High Performance	38%	21%	41%
Medium Performance	34%	21%	45%
Low Performance	24%	20%	56%

The maximum number of hostiles that could have been destroyed in the low workload experiment was 2, while a total of 5 hostiles could have been destroyed in the moderate workload experiment. As shown in the low workload data, 5 participants destroyed 0 hostiles during the 4-hour mission. However, in the moderate workload data, the worst 4 performers destroyed one hostile. On the other hand, no participants in the moderate workload experiment were able to destroy all 5 available hostiles, but over one third of participants in the low workload study were able to achieve the mission objective of destroying all hostiles.

As seen in both the search and destroy data sets, participants can achieve the highest performance as well as the lowest performance under long duration, low workload conditions of the multiple UxV supervisory control scenario. The moderate workload environment appears more predictable, but compared to the low workload environment, neither the best nor the worst possible performance is achieved.

This comparison between workload levels and performance brings this research discussion full circle, back to the first research question of whether the Yerkes-Dodson curve holds true for low workload. It can be seen that, while the worst possible performance can occur during low workload, that is not as likely. Therefore, according to this research, the parabolic drop in performance at low workload suggested by the Yerkes-Dodson curve was not confirmed as the model for how operators perform in a low workload, supervisory control environment. Perhaps the automation made up for times when the participants could not focus on the mission, and the distractedness of the participants actually helped sustain alertness. The majority of the data showed that mediocre and even exemplary performance can be achieved at low workload. However, this is not to say the participants enjoyed the low workload environment. Their survey comments and pained looks in the video data demonstrated the extreme boredom and tedious nature associated with the low workload environment. Despite the hardships of the long duration, low workload experiment, one third of participants still exceeded the performance of the moderate workload experiment in destroying hostiles. This research finding suggests that excellent performance can be achieved amid tedious conditions of long duration, low workload missions.

5 Conclusion

This research revealed that a low workload environment for supervisory control of decentralized heterogeneous unmanned vehicles impacts operators' vigilance and attention state. This experiment provided a unique environment for participants to perform a complex supervisory control task while allowing them to react to the boredom environment in their own way. This research was able to simultaneously gather objective performance data in a realistic search, track, and destroy UxV mission and capture the natural boredom behaviors induced by the grueling simulation. Humans have to employ coping mechanisms to surmount the boredom of prolonged low workload. Low workload has a way of bringing out the best performance in people, while bringing out the worst in others.

This research determined that the Yerkes-Dodson curve, which predicts that performance plummets at low workload, does not hold true for low workload in supervisory control of networked UxVs. People subjected to low workload can perform equally well if not better than operators working at moderate workload.

This researched also uncovered results that were not foreseen. Incrementing lower levels of taskload does not necessarily decrease operator utilization, or percent busy time. This experiment discovered that participants self-imposed interactions with the human-computer system when subjected to a low taskload scenario. Under these conditions, operators displayed directed attention toward their assigned work only a third of the time. Moreover, the operators hardly multitasked, perhaps because dividing their attention requires the extra effort of doing more than one thing at once. This low workload environment caused vigilance to degrade over time, as shown by the decreasing directed attention, especially during the second half of the mission.

This research brought to light key characteristics that can predict performance in a prolonged supervisory control mission under low workload. Video gamers are predicted to be poor performers in a low workload supervisory control environment because they are conditioned to the need for constant stimuli. In a long duration, low workload mission, increasing utilization predicts better performance in the search and destroy tasks of supervisory control of networked UxVs. High directed attention can predict good mission performance, even in the track task, which is mainly automated.

Lastly, this research provides evidence contrary to the common belief that distraction is harmful to mission performance. It was shown that the majority of the top performers had a high percentage of distraction time. Distraction can be a method for keeping the mind and body engaged and alert. When used in conjunction with effective switching times, distraction can help operators attain top performance.

5.1 Possible Solutions

The concept of automated adaptation can be considered a solution to the detriments of low workload. It has been shown that implementing certain automation adaptation with certain levels of operator workload enhances performance [67].

Adaptive aiding can be implemented in times of high operator workload to help the operator cope with high workload. On the other hand, adaptive task allocation can also be implemented at low levels of operator workload for the purpose of bringing the operator up to a comfortable workload in order to improve performance [67]. Adaptive automation may help mitigate the harmful effects of low workload discovered in this study, but more research is needed to determine how to use effective adaptive techniques.

5.2 Additional Future Work

A high workload experiment could be conducted to add to the low workload and moderate workload studies previously discussed. In that way, a full range of performance data spanning low, moderate, and high workload could be plotted to make a complete assessment of the Yerkes-Dodson relationship of performance to workload.

Future work can also be conducted to model human interaction with multiple UxVs in low workload conditions. The goal would be to have a model that accounts for boredom and spikes in workload in order to predict operator performance. Switching time research needs to be conducted in order to implement the performance aid of switching times into the human performance model. This future research will assist in the design of smart decision support tools that can increase vigilance and performance of operators in supervisory control domains with low workload. The research of this thesis paves the way for future research on modeling boredom in supervisory control of networked UxVs.

Appendix A: Interface Details

This Appendix describes the OPS-USERS interface.

A.1 UxV Symbols

The UxV symbols displayed in the map view are depicted in the following table.

Table 7: UxV Symbols

	Vehicle Type	Range and Fuel	Radar Footprint	Primary Mission	Image
USV 1 Unmanned Surface Vehicle	Ship that runs along the river	Medium	Large	Search and Track	
UAV 2 Unmanned Aerial Vehicle	Fixed-wing airplane	Small	Rectangular due to mounted camera	Search and Track	
UAV 3 Unmanned Aerial Vehicle	Helicopter	Small	Rectangular due to mounted camera	Search and Track	
WUAV Weaponized Unmanned Aerial Vehicle	Helicopter	Large	Large	Detect and Destroy Hostiles	

A.2 Refueling Base

The UxVs refuel themselves automatically at the base location symbol.

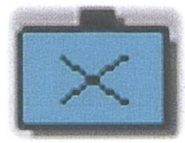

Figure 22: Refueling Base

A.3 Search Task Symbols

Search tasks can be added to the mission. A "search" task designates a location for a UxV to go to in search of a target.

- Color shows priority level.
- The letter to the right of the search task identifies it (this is its name).
- The number to the left of the search task symbol indicates which UxV is assigned to perform the search task (note than search task F is unassigned).

Figure 23: Search Task Symbols

For example, the search task on the left is called search task D. UAV 3 is assigned to travel to the location on the map where this search task symbol resides. UAV 3 will search the area at the search task location and during the transit to the location.

A.4 Target Symbols

The UxVs must periodically track or revisit the targets that have been found. The Weaponized UAV must destroy hostile targets. The shape and color of the target symbols is a dual coding of their representation to benefit colorblind operators.

- Red diamonds are hostile targets.
- Yellow clovers are unknown targets.
- Blue rectangles are friendlies and are not tracked.
- The letter on the right identifies the target.
- The character on the left indicates which UxV is assigned to the target (for example, the Weaponized UAV is assigned to destroy hostile target D shown in Figure 24).

Figure 24: Target Symbols

According to the center symbol, UAV 2 will track Unknown Target B. UAV 2 will travel to the location where this target symbol is positioned on the map and begin following the target. If UAV 2 has another task to perform or must go back to base to refuel, the computer algorithm will calculate an estimated new position for the target based on the target's last known position and velocity.

Flags attached to the target symbols designate priority level. The color and location of the flag is a dual coding of its representation to benefit colorblind operators.

- Red vertical flag on top of the target symbol specifies high priority.
- Orange horizontal flag beside the target symbol specifies medium priority.
- Yellow downward flag below the target symbol specifies low priority.
- Friendlies do not have a priority level flag because they do not need to be tracked.
- Figure 25 shows some example priority level-designated targets.

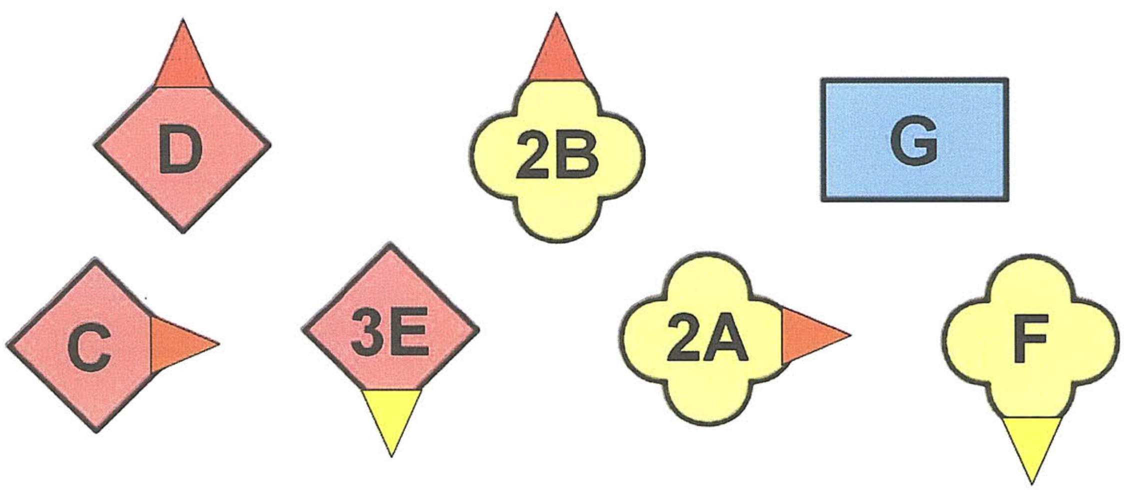

Figure 25: Target Priority Flags

A.5 Loiter Symbols

The Weaponized UAV does not search or track targets. The WUAV can only detect targets and destroy hostile targets. The WUAV can be sent to loiter, or hover over a particular position, while waiting to destroy hostile targets. The loiter symbol for the WUAV resembles a stop sign. The color indicates priority level.

Figure 26: Loiter Symbols

A.6 Target Identification Sequence

Initially the target symbol may not be visible within the target identification window. The participant must click and drag over the area within the window to pan for the target symbol. Subsequently, the participant can click the appropriate target designation button to identify the target symbol as hostile, unknown, or friendly. If an unknown target is found, the target must first be marked as unknown. However, the designation can be edited later as more information arises from the chat box. Once the target has been identified, the system allows the participant to choose a priority level for the emergent target. The command center provides information on the priority levels of emergent targets based on the location of target discovery. This priority level information is disseminated via the chat message box. Figure 27 depicts this sequence of target finding, panning to observe the target symbol, identifying the target, and designating a priority level.

Figure 27: Target Identification Sequence

A.7 Destroyed Hostiles

Destroyed targets appear as black symbols on the Map View. These destroyed target symbols remain on the map for the duration of the simulation to indicate the destruction sites.

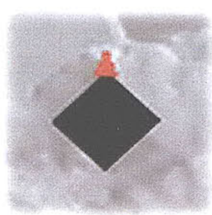

Figure 28: Destroyed Hostile Target Symbol

Appendix B: Consent to Participate Form

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Long Duration, Low Workload Missions for Heterogenous Unmanned Vehicle Teams

You are asked to participate in a research study conducted by Professor Mary Cummings PhD, from the Aeronautics and Astronautics Department at the Massachusetts Institute of Technology (M.I.T.). You were selected as a possible participant in this study because the expected population this research will influence is expected to contain men and women between the ages of 18 and 50 with an interest in using computers. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

The purpose of this research is to see what the effect is of a long duration, low workload scenario in the context of piloting multiple, highly autonomous, unmanned vehicles in the setting of a populated control room.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

- Participate in training on the video game-like interface via the refresher tutorial, as you are already familiar with the interface from the previous OPS-USERS experiment. Complete a fifteen-minute practice session where control a team of simulated unmanned vehicles. The vehicles you will control will be assigned with the task of finding, identifying, and tracking targets in an area of interest, destroying hostile targets, and collaborating with the auto-planner to replan schedules.

- Participate in a four-hour long testing session where you will experience a long duration, low workload mission. You will work alongside two other participants to simulate a populated control room, and you will each have your own workstations with your own vehicles and territory to control
- You will be rewarded a score for the trial based on the number of targets you successfully find, how long they are successfully tracked thereafter, the percentage of the total area of interest is searched, and number of hostile targets destroyed.
- All testing will take place at MIT in room 35-220.
- Total time: 4 hours and 45 minutes

- **POTENTIAL RISKS AND DISCOMFORTS**

There are no anticipated physical or psychological risks in this study.

- **POTENTIAL BENEFITS**

While you will not directly benefit from this study, the results from this study will assist in the design of interfaces for human-UV systems.

- **PAYMENT FOR PARTICIPATION**

You will be paid $125 to participate in this study which will be paid upon completion of your debrief. Should you elect to withdraw in the middle of the study, you will be compensated for the hours you spent in the study. An additional $250 Best Buy Gift Card will be awarded to the participant with the high score.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. You will be assigned a subject number which will be used on all related documents to include databases, summaries of results, etc.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact the Principal Investigator, Mary L. Cummings, at (617) 252-1512, e-mail, missyc@mit.edu, and her address is 77 Massachusetts Avenue, Room 33-311, Cambridge, MA, 02139. The investigators are Christin Hart and Vicki Crosson. They may be contacted at (617) 253-0993 or via email at chart@mit.edu and viccro@mit.edu.

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____

Name of Subject

____

Name of Legal Representative (if applicable)

Signature of Subject or Legal Representative Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

Signature of Investigator Date

Appendix C: Demographic Survey

1. Subject number:_____

2. Age:_____

3. Gender: *M* *F*

4. Occupation:______________________________

   if student, (circle one): *Undergrad* *Masters* *PhD*

   expected year of graduation:_________

5. Military experience (circle one): *No* *Yes* If yes, which branch:________

   Years of service:________

6. Give an overall rating of your past two nights of sleep.

   Poor *Fair* *Good* *Great*

7. Rate your health in terms of nutrition and exercise in the past week.

   Poor *Moderate* *Good*

8. How often do you play computer games?

   Rarely *Monthly* *Weekly* *A few times a week* *Daily*

   Types of games played:______________________________________

9. Rate your comfort level with using computer programs.

   Not comfortable *Somewhat comfortable* *Comfortable* *Very Comfortable*

10. What is your perception toward unmanned vehicles?

    Intense dislike *Dislike* *Neutral* *Like* *Really Like*

Appendix D: Demographic Results

In a demographic survey, participants were asked to rate their gaming experience, computer comfort level, and perception toward unmanned vehicles. The demographic survey can be found in Appendix C. Participants indicated their frequency of playing video games on a five-point Likert scale from "rarely plays games" to "daily gamer." Participants can essentially be grouped into two video gaming categories: gamers and non-gamers, where gamers played at least weekly and non-gamers only played games monthly or rarely. Thus, one third of participants were gamers and two thirds were non-gamers. Table 8: Gaming Demographics shows the category of gamer versus non-gamer associated with each level of gaming frequency in addition to the number of participants who indicated that Likert scale level.

Table 8: Gaming Demographics

Gaming Frequency	**Rarely**	**Monthly**	**Weekly**	**Multi-weekly**	**Daily**
Gaming Level	Non-gamer	Non-gamer	Gamer	Gamer	Gamer
Number of Participants	11	9	7	3	0

The computer comfort level 4-point Likert scale rating ranges from not comfortable to very comfortable. The vast majority of participants indicated a high comfort level with using computer programs, as shown in Table 9.

Table 9: Computer Comfort Level Demographics

Computer Comfort Level	**Not Comfortable**	**Somewhat Comfortable**	**Comfortable**	**Very Comfortable**
Number of Participants	1	4	12	13

The five-point Likert scale for perception toward unmanned vehicles ranges from "intense dislike" to "really like" with a neutral category in the middle. Overall participants either liked unmanned vehicles or felt neutral; these demographics on UxV perception show a shift since the previous experiment with a very similar pool of subjects (some of whom changed their mind about UxVs). These results are shown in Table 10.

Table 10: Perception Toward UxVs Demographics

Perception toward UxVs	Intense Dislike	Dislike	Neutral	Like	Really Like
Moderate Workload Study	0	1	37	43	17
Low Workload Study	3	0	20	8	0

Appendix E: Pre-experiment Skill Survey

1. How confident were you about the actions you took?

   Not Confident   Somewhat Confident   Confident   Very Confident   Extremely Confident

2. How did you feel you performed?

   Very Poor   Poor   Satisfactory   Good   Excellent

3. How busy did you feel during the practice mission?

   Extremely Busy   Busy   Not Busy   Idle

4. Do you understand how to create search tasks?

   No   Somewhat   Yes

5. Do you understand how to use the target identification window?

   No   Somewhat   Yes

6. Do you understand how to approve a weapon launch on hostile targets?

   No   Somewhat   Yes

7. Do you understand how to use the Schedule Comparison Tool (SCT)?

   No   Somewhat   Yes

8. Do you understand that you must accept a plan in order for the unmanned vehicles to perform new search, track and destroy tasks?

   No   Somewhat   Yes

9. Do you understand that, while in the Schedule Comparison Tool, you have the option to cancel without accepting a plan?

   No   Somewhat   Yes

Appendix F: Post-experiment Survey

1. How confident were you about the actions you took?

   Not Confident   Somewhat Confident   Confident   Very Confident   Extremely Confident

   Comments:

2. How did you feel you performed?

   Very Poor   Poor   Satisfactory   Good   Excellent

3. How busy did you feel during the mission?

   Idle   Not Busy   Busy   Very Busy   Extremely Busy

4. Did you feel distracted?   *Yes   No*

   If so, please list some of the items or activities that distracted you from the mission:

5. Other comments:

Appendix G: Linear Regression Coeffi cient Tables

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Target Finding Score	1.891	2	24	.173
Hostile Destruction Score	.861	2	24	.435
Target Tracking Percentage	4.063	2	25	.030

Tests of Normality

	Kolmogorov-Smirnov(a)		
	Statistic	df	Sig.
Target Finding Score	.154	26	.116
Hostile Destruction Score	.188	26	.019
Target Tracking Percentage	.319	26	.000

G.1 Target Finding Score

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	.530	.252		2.099	.048
	Total Directed	.330	.366	.183	.902	.377
	Total Divided	.395	.394	.172	1.004	.326
	Total UT	-4.920	1.644	-.579	-2.992	.007
	Gaming	.184	.105	.319	1.759	.092
2	(Constant)	.601	.239		2.519	.019
	Total Divided	.415	.392	.181	1.059	.301
	Total UT	-4.206	1.435	-.495	-2.931	.008
	Gaming	.153	.098	.264	1.553	.134
3	(Constant)	.715	.214		3.341	.003
	Total UT	-4.275	1.437	-.504	-2.975	.007
	Gaming	.139	.098	.241	1.424	.167
4	(Constant)	.906	.170		5.341	.000
	Total UT	-4.282	1.466	-.504	-2.920	.007

Dependent Variable: Target Finding Score

G.2 Target Tracking Percentage

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	.998	.031		31.858	.000
	Total Directed	.131	.063	.427	2.074	.049
	Total UT	-.637	.307	-.427	-2.078	.048

Dependent Variable: Target Tracking Percentage

G.3 Hostile Destruction Score

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	1.227	.635		1.933	.066
	Total_Focused	-.296	.920	-.067	-.322	.750
	Total_Divided	.068	.990	.012	.069	.946
	TotalUT	-8.405	4.136	-.405	-2.032	.054
	Gaming	.491	.264	.348	1.863	.076
2	(Constant)	1.245	.567		2.196	.038
	Total_Focused	-.293	.899	-.067	-.326	.747
	TotalUT	-8.424	4.037	-.406	-2.087	.048
	Gaming	.489	.256	.347	1.908	.069
3	(Constant)	1.177	.517		2.276	.032
	TotalUT	-9.055	3.475	-.437	-2.606	.015
	Gaming	.518	.236	.367	2.190	.038

a Dependent Variable: HostileDestructionScore

Appendix H: Hourly Pairwise Comparisons

Pairwise Comparisons

Measure: MEASURE_1

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	.129*	.030	.000	.067	.192
	3	.195*	.032	.000	.130	.261
	4	.175*	.028	.000	.117	.233
2	1	-.129*	.030	.000	-.192	-.067
	3	.066*	.023	.008	.019	.113
	4	.046	.024	.066	-.003	.095
3	1	-.195*	.032	.000	-.261	-.130
	2	-.066*	.023	.008	-.113	-.019
	4	-.020	.020	.327	-.061	.021
4	1	-.175*	.028	.000	-.233	-.117
	2	-.046	.024	.066	-.095	.003
	3	.020	.020	.327	-.021	.061

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Factor comparisons represent the four hour mission duration: hours 1, 2, 3, and 4.

Appendix I: Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Required Utilization	30	.01	.03	.02	.00
Self Imposed Utilization	30	.03	.15	.10	.03
Total Utilization	30	.05	.18	.11	.03
Performance Score	27	.23	5.3	2.2	1.6
Target Finding Score	27	.12	1.3	.43	.28
Hostile Destruction Score	27	.01	2.0	.88	.69
Target Tracking Percent	28	.80	1.0	.97	.05
Number of Search Tasks Created	30	57	340	190	68.
Replan Avg Reaction Time	27	1.7	27	8.6	7.1
Prompted Search Avg Reaction Time	30	10.	30.	21.	6.9
Chat Avg Reaction Time	28	3.0	48	19	11
Chat Accuracy	30	.33	1.0	.89	.20
Prompted Search Task Accuracy	30	.25	1.0	.73	.23
Extra Search Tasks	30	42	330	180	68
Extra Replans	30	46	370	190	74
Extra Target Edits	29	.00	12	4.0	3.6
Number of Targets Lost & Found	29	.00	4.0	.93	1.2
Age	30	19	32	23.	3.0
Sleep Self Rating	29	1	4	2.6	.78
Health Self Rating	30	1	3	2.6	.57
Gaming Level	30	1	2	1.3	.48
Gaming Experience	30	1	4	2.1	1.0
Computer Comfort Level	30	1	4	3.2	.81
UxV Perception	30	2	5	3.7	.79
Self Rated Confidence	30	3	5	37	.55
Self Rated Performance	30	2	5	3.7	.61
Self Rated Busyness	30	1.0	3.0	1.9	.56
Divided SociallyDirected with poor posture	30	.00	.10	.03	.03
Divided PhysicallyDivided Socially	30	.03	.55	.17	.13
Divided CognitivelyDivided Physically	29	.00	.13	.01	.03
Distracted SociallyDivided Cognitively	30	.00	.29	.09	.09
Distracted PhysicallyDistracted Socially	30	.00	.18	.06	.05
Distracted CognitivelyDistracted Physically	30	.04	.59	.29	.15
Total DirectedDistracted Cognitively	30	.10	.75	.34	.15
Total DividedTotal Directed	30	.09	.55	.22	.13
Total DistractedTotal Divided	30	.07	.79	.44	.20

Appendix J: Sources of Error

If this experiment were to be repeated, certain aspects of the study could be controlled better. Perhaps a psychological profile could be conducted before the experiment to cross-reference personalities with boredom coping mechanisms. Video footage that simultaneously shows a clear close-up of each operator's face as well as the distraction material they are engaging could result in more accurate video coding. One video source served as the footage for all three participants in each test session, and a clearer view of each participant and their surroundings could be attained with separate cameras focusing on each participant.

A more stable simulation would improve the testing environment. Nine of 39 participants' data had to be discarded because of simulation crashes, and the system failures interrupted the test session each time. In addition, more controlled movement of the hidden targets could have been achieved to ensure all participant scenarios were equivalent in terms of hidden hostile targets uncloaking and quantity. A more robust automated planner would remove participant frustration with the automation and make for a more controlled study. All of these sources of error could be accounted for in order to improve the validity of independently verified results.

Appendix K: Top Performer Demographics

Rank	Age	Sleep Self-rating	Health Self-rating	Computer Comfort Level	UxV Perception
1	19	2	2	2	3
2	28	3	2	3	5
3	23	3	3	1	3
4	23	2	3	3	4
5	23	4	2	2	3
6	23	3	3	3	4
7	23	3	3	4	5
8	23	3	2	3	4

Rank	Confidence Self-rating	Performance Self-rating	Busyness Self-rating
1	3	4	3
2	4	3	2
3	4	4	3
4	3	4	2
5	4	4	1
6	4	4	2
7	4	5	2
8	3	4	2

References

[1] C.E. Nehme, J.W. Crandall, and M.L. Cummings, "An Operator Function Taxonomy for Unmanned Aerial Vehicle Missions," in *12th International Command and Control Research and Technology Symposium*, Newport, Rhode Island, 2007.

[2] U.S. Navy, "Dragon Warrior Communications Relay," http://cs.itd.nrl.navy.mil/work/dragon_warrior/index.php (accessed January 13, 2010).

[3] P. Smith, E. McCoy, and C. Layton, "Brittleness in the Design of Cooperative Problem-solving Systems: The Effects on User Performance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 27, no. 3, pp. 360-371, 1997.

[4] T.B. Sheridan, *Telerobotics, Automation and Human Supervisory Control*. Cambridge, MA: The MIT Press, 1992.

[5] A.P. Nicholls, A. Melia, E.W. Farmer, G. Shaw, T. Milne, A.W. Stedmon, S. Sharples, and G. Cox, "Changing the Role of the Air Traffic Controller: How Will Free Flight Affect Memory for Spatial Events?," *Applied Ergonomics*, vol. 38, no. 4, pp. 457-463, 2007.

[6] R. N. Charette, "Automated to Death," http://spectrum.ieee.org/computing/software/automated-to-death/0 (accessed January 13, 2010).

[7] R.I. Thackray, *Boredom and Monotony as a Consequence of Automation: A Consideration of the Evidence Relating Boredom and Monotony to Stress*, DOT/FAA/AM-80/1 ed. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1980.

[8] J. Langan-Fox, M.J. Sankey, and J.M. Canty, "Keeping the Human in the Loop: From ATCOs to ATMs," in *Keynote Speech by J. Langan-Fox at the Smart Decision Making for Clean Skies Conference*, Canberra, Australia, July 2008.

[9] Naval Studies Board, "Autonomous Vehicles in Support of Naval Operations," National Research Council, Washington DC, 2005.

[10] M.L. Cummings, S. Bruni, S. Mercier, P.J. Mitchell, "Automation Architecture for Single Operator, Multiple UAV Command and Control," *The International Command and Control Journal*, vol. 1, no.2, pp. 1-24, 2007.

[11] A. Reiner, "Pilots on Autopilot," http://www.nytimes.com/2009/12/17/opinion/17reiner.html_r=1 (accessed December 16, 2009).

[12] J.S. Warm, R. Parasuraman, and G. Mattews, "Vigilance Requires Hard Mental Work and is Stressful," *Human Factors*, vol. 50, no. 3, pp. 433-441, 2008.

[13] V.D. Hopkin, *Human Factors in Air Traffic Control*. London: Taylor and Francis, 1995.

[14] S. Straussberger, "Monotony in ATC: Contributing Factors and Mitigation Strategies," Ph.D. Dissertation. Karl-Franzens University, 2006.

[15] C.D. Wickens, A.S. Mavor, and J.P. McGee, *Flight to the Future: Human Factors in Air Traffic Control*. Washington DC: National Academy Press, 1997.

[16] S. Straussberger and D. Schaefer, "Monotony in Air Traffic Control," *Air Traffic Control Quarterly, International Journal of Engineering and Operations,* vol. 15, no. 3, pp. 183-207, 2007.

[17] V.J. Davies, D.R. Shackleton, and R. Parasuraman, "Monotony and Boredom," in *Stress and Fatigue in Human Performance,* G.R.J. Hockey, Ed. New York: Wiley, 1983, pp. 1-32.

[18] R.I. Thackray, "The Stress of Boredom and Monotony," *Psycosomatic Medicine,* vol. 43, no. 2, pp. 165-176, 1981.

[19] C. Wickens and J.G. Hollands, *Engineering Psychology and Human Performance.* Upper Saddle River, N.J.: Prentice Hall, 2000.

[20] R.M. Yerkes and J.D. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *Journal of Comparative Neurology and Psychology,* vol. 18, no. 5, pp. 459-482, 1908.

[21] C.E. Nehme, "Modeling Human Supervisory Control in Heterogeneous Unmanned Vehicle Systems," in *Aeronautics and Astronautics*. Ph.D Dissertation. Massachusetts Institute of Technology, 2009.

[22] P.A. Hancock, J.S. Warm, "A Dynamic Model of Stress and Sustained Attention," *Human Factors and Ergonomics Society*, vol. 31, no. 5, pp. 519-537, 1989.

[23] D.O. Hebb, "Drives and the CNS," *Psychological Review,* vol. 62, pp. 243-254, 1955.

[24] M.L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors,* vol. 49, no.1, pp. 1-15, 2007.

[25] M.L. Cummings, A.S. Brzezinski, and J.D. Lee, "The Impact of Intelligent Aiding for Multiple Unmanned Aerial Vehicle Schedule Management," *IEEE Intelligent Systems: Special Issue on Interacting with Autonomy,* vol. 22, no.2, pp. 52-59, 2007.

[26] A.S. Clare, C.S. Hart, M.L. Cummings, "Assessing Operator Workload and Performance in Expeditionary Multiple Unmanned Vehicle Control," in *48th American Institute of Aeronautics and Astronautics Aerospace Sciences Meeting,* Orlando, Florida, 2009.

[27] D.K. Schmidt, "A Queuing Analysis of the Air Traffic Controller's Workload," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 8, no. 6, pp. 492-498, 1978.

[28] W.B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.

[29] B. Donmez, C. Nehme, M.L. Cummings, and P. de Jong, "Modeling Situational Awareness in Multiple Unmanned Vehicle Supervisory Control Discrete Event

Simulation," *IEEE Intelligent Systems: Man and Cybernetics, Part A Systems and Humans*.
[30] P.A. Hancock, "Arousal Theory, Stress, and Performance: Problems of Incorporating Energetic Aspects of Behavior into Human-machine System Function," in *Ergonomics and Human Factors: Recent Research*, J.S. Warm, L.S. Mark, and R.L. Huston, Ed., New York: Springer-Verlag, 1987, pp. 170-179.
[31] G.R.J. Hockey and P.A. Hamilton, "The Cognitive Patterning of Stress States," in *Stress and Fatigue in Human Performance*, G.R.J. Hockey, Ed., New York: Wiley, 1983, pp. 331-361.
[32] R. Naatanen, "The Inverted-U relationship between activation and performance: A Critical Review," in *Attention and Performance IV*, S. Kornblum, Ed. New York: Academic, 1973.
[33] S.R. Schneider and R.M. Shiffrin, "Controlled and Automatic Human Information-Processing," *Psychological Review*, vol. 1, pp. 1-66, 1977.
[34] A.D. Fisk and W. Schneider, "Control and Automatic Processing During Tasks Requiring Sustained Attention: A New Approach to Vigilance," *Human Factors*, vol. 23, no. 6, pp. 737-750, 1981.
[35] R.A. Grier, J.S. Warm, W.N. Dember, G. Matthews, T.L Galinksky, and J.L. Szalma, "The Vigilance Decrement Reflects Limitations in Effortful Attention, Not Mindlessness," *Human Factors*, vol. 45, no. 3, pp. 349-359, 2003.
[36] J.G. Temple, J.S. Warm, W.N. Dember, K.T.S. Jones, C.M. LaGrange, and G. Matthews, "The Effects of Signal Salience and Caffeine on Performance, Workload, and Stress in an Abbreviated Vigilance Task," *Human Factors*, vol. 42, no.2, pp. 183-194, 2000.
[37] W.S. Helton, T.D. Hollnader, J.S. Warm, G. Matthews, W.N. Dember, M. Wallart, G. Beauchamp, R. Parasuraman, and P. Hancock, "Signal Regularity and the Mindlessness Model of Vigilance," *British Journal of Psychology*, vol. 96, no. 2, pp. 249-261, 2005.
[38] T. Manly, I.H. Robertson, M. Galloway, and K. Hawkins, "The Absent Mind: Further Investigations of Sustained Attention to Response," *Neruopsychologia*, vol. 37, no. 6, pp. 661-670, 1999.
[39] I.H. Robertson, T. Manly, J. Andrade, B.T. Baddeley, and J. Yiend, "Oops!: Performance Correlates of Everyday Attentional Failures in Traumatic Brain Injured and Normal Subjects" *Neruopsychologia*, vol. 35, no. 6, pp. 747-758, 1997.
[40] D.T. Struss, T. Shallice, M.P. Alexander, and T.W. Picton, "A Multidisciplinary Approach to Anterior Attentional Functions," in *Ann NY Academy Science*, vol. 769, no. 1, pp. 191-209, 1995.
[41] J. Langan-Fox, M.J. Sankey, and J.M. Canty, "Human Factors Measurement for Future Air Traffic Control Systems," *Human Factors*, vol. 51, no. 5, pp. 595-637, 2009.

[42] B.T. Stollery, "Vigilance," in *International Encyclopedia of Ergonomics and Human Factors*, vol. 1, W. Karwowski, Ed., Boca Raton, FL: CRC Press, pp. 965-968, 2006.
[43] S. Makeig, F.S. Elliot, M. Inlow, and D.A. Kobus, "Lapses in Alertness: Brain-evoked Responses to Task-irrelevant Auditory Probes," Report No. 90-39, Naval Heath Research Center, San Diego, CA,1992.
[44] D.J. Schroeder, R.M. Touchston, J.A. Stern, N. Stoliarov, and R. Thackray, *Maintaining Vigilance on a Simulated ATC Monitoring Task Across Repeated Sessions*. Oklahoma City: Federation Aviation Administration, Civil Aeromedical Institute, 1994.
[45] R.I. Thackray and R.M. Touchston, *An Evaluation of the Effects of High Visual Taskload on the Separate Behaviors Involved in Complex Monitoring Performance*. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1988.
[46] T.H. Shaw, J.S. Warm, V. Finomore, L. Tripp, G. Matthews, E. Weiler, and R. Parasuraman, "Effects of Sensory Modality on Cerebral Blood Flow Velocity During Vigilance," *Neuroscience Letters*, vol. 461, no. 3, pp. 207-211, 2009.
[47] T.-P. Jung, S. Makeig, M. Stensmo, and T.J. Sejnowski, "Estimating Alertness from the EEG Power Spectrum," *IEEE Transactions on Biomedical Engineering,* vol. 44, no. 1, pp. 60-69, 1997.
[48] A.J. Rehmann, *Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research*. Atlantic City, NJ: Federal Aviation Administration, 1995.
[49] N. Pattyn, X. Neyt, D. Henderickx, and E. Soetens, "Psychophysiological Investigation of Vigilance Decrement: Boredom or Cognitive Fatigue?," *Physiology and Behavior,* vol. 93, no. 1-2, pp. 369-378, 2008.
[50] P. Stager, D. Hameluck, and R. Jubis, "Underlying Factors in Air Traffic Control Incidents," in *Proceedings of the Human Factors Society 33rd Annual Meeting,* Santa Monica, CA, pp. 43-46, 1989.
[51] A.B. Hill and R.E. Perkins, "Towards a Model of Boredom," *British Journal of Psychology,* vol. 76, no.2, pp. 235-240, 1985.
[52] R.I. Thackray, J. Powell, M.S. Bailey, and R.M. Touchstone, *Physiological, Subjective, and Performance Correlates of Reported Boredom and Monotony while Performing a Simulated Radar Control Task*. Oklahoma City: Federal Aviation Administration, Civil Aeromedical Institute, 1975.
[53] S.J. Kass, S.J. Vodanovich, C. Stanny, and T.M. Taylor, "Watching the Clock: Boredom and Vigilance Performance," *Perceptual Motor Skills,* vol. 92, no. 3, pp. 969-976, 2001.
[54] M.D. Rodgers and L.G. Nye, "Factors Associated with Severity of Operational Errors at Air Route Traffic Control Centers," in *An Examination of the Operational Error Database for Air Traffic Control Centers*, M. D. Rodgers, Ed., Washington DC: Federal Aviation Administration, Office of Aviation Medicine, pp. 243-256, 1993.

[55] G. Costa, *Occupational Stress and Stress Prevention in Air Traffic Control*, Geneva, Switzerland: Conditions of Work and Welfare, International Labor Office, 1995.

[56] M. Jacobs, B. Fransen, J. McCurry, F.W.P. Heckel, A.R. Wagner, and J. Trafton, "A Preliminary System for Recognizing Boredom," in *ACM/IEEE International Conference on Human-Robot Interaction* La Jolla, CA, pp. 299-300, 2009.

[57] S. D'Mello, P. Chapman, and A. Graesser, "Posture as a Predictor of Learner's Affective Engagement," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, Austin TX, pp. 905-910, 2007.

[58] D.A. Sawin and M.W. Scerbo, "Effects of Instruction Type and Boredom Proneness in Vigilance: Implications for Boredom and Workload," *Human Factors*, vol. 37, no. 4, pp. 752-765, 1995.

[59] K. Button, "Different Courses: New Style UAV Trainees Edge Toward Combat," http://www.c4isrjournal.com/story.php?F=4322170 (accessed on January 16, 2010).

[60] W.T. Thompson, N. Lopez, P. Hickey, C. DaLuz, and J.L. Caldwell, "Effects of Shift Work and Sustained Operations: Operator Performance in Remotely Piloted Aircraft (OP-REPAIR)," 311th Performance Enhancement Directorate and Performance Enhancement Research Division, Brooks City-Base, Texas, 2006.

[61] S. Fisher, "Replan Understanding for Heterogenous Unmanned Vehicle Teams," Master's Thesis, Massachusetts Institute of Technology, 2008.

[62] U.S. Department of Defense, "Interface Standard, Common Warfighting Symbology, Report number: MIL-STD-2525B," 1999.

[63] M.L. Cummings, A. Clare, and C. Hart, "The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling," *Human Factors and Ergonomics Society*, vol. 52, no. 2, 2010.

[64] M.L. Cummings and C.E. Nehme, "Modeling the Impact of Workload in Network Centric Supervisory Control Settings," in *Neurocognitive and Physiological Factors During High-Tempo Operations*, S. Kornguth, M.D. Matthews, and R. Steinberg, Ed., Ashgate, 2010.

[65] B. Donmez, P.E. Pina, and M.L. Cummings, "Evaluation Criteria for Human-Automation Performance Metrics," *Paper presented at the Performance Metrics for Intelligent Systems Workshop*, Gaithersburg, MD, 2008.

[66] M.S. Young and N.A. Stanton, "Malleable Attention Resources Theory: A New Explanation for the Effects of Mental Underload on Performance," *Human Factors*, vol. 44, no. 3, pp. 365-375, 2002.

[67] R. Parasuraman, M. Mouloua, and B. Hilburn, "Adaptive Aiding and Adaptive Task Allocation Enhance Human-Machine Interaction," in *Automation Technology and Human Performance*, M.W. Scerbo and M. Mouloua, Ed., New Jersey: Lawrence Erlbaum Associates, Inc., pp. 119-123, 1999.