Summer 8-2020

Exploring the Effects of Task Priority on Attention Allocation and Trust Towards Imperfect Automation: A Flight Simulator Study

Tetsuya Sato

Old Dominion University, sato.tetsuya1996@gmail.com

EXPLORING THE EFFECTS OF TASK PRIORITY ON ATTENTION ALLOCATION AND TRUST TOWARDS IMPERFECT AUTOMATION: A FLIGHT SIMULATOR STUDY

by

Tetsuya Sato
B.S. May 2018, Old Dominion University

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

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
August 2020

Approved by:

Yusuke Yamani, Ph.D. (Director)

Mark W. Scerbo, Ph.D. (Member)

Jeremiah D. Still, Ph.D. (Member)

ABSTRACT

EXPLORING THE EFFECTS OF TASK PRIORITY ON ATTENTION ALLOCATION AND TRUST TOWARDS IMPERFECT AUTOMATION: A FLIGHT SIMULATOR STUDY

Tetsuya Sato
Old Dominion University, 2020
Director: Dr. Yusuke Yamani

The present study examined the effect of task priority and task load on attention allocation and automation trust in a multitasking flight simulator platform. Previous research demonstrated that, participants made less fixations and reported lower levels of trust towards the automation in the secondary monitoring under higher load on the primary tracking task (e.g., Karpinsky et al., 2018). The results suggested that participants perceived behaviors of the automated system less accurately due to less attention allocated to monitoring of the system, leading to decreased trust towards it. One potential explanation of the effect is that participants might have prioritized the tracking task due to the elevated task load over monitoring of the automation. The current study employed a 2 x 2 mixed design with task difficulty (low vs. high difficulty) and task priority (equal vs. tracking priority). Participants performed the central tracking task, the system monitoring task, and the fuel management task where the system monitoring was assisted by an imperfect automated system. Participants were instructed to either prioritize the central tracking task over the other two tasks or maximize performance for all tasks. Additionally, participants received feedback on their tracking performance reflecting an anchor of their baseline performance. The data indicated that participants rated lower performance-based trust in a multitasking environment when all tasks were equally prioritized, supporting the notion that task priority modulates the effect of task load.

This thesis is dedicated to Mom, Dad, Grandma, Grandpa, and my other family members.

Without their support, I would not have made it this far.

ACKNOWLEDGEMENTS

There are several people that I would like to acknowledge for helping me complete my thesis. I would like to express my greatest appreciation to Dr. Yusuke Yamani for his endless support throughout my graduate year. I would not have accomplished this big milestone without his guidance and support. I am grateful that I was able to have him as my academic advisor. Also, I would like to thank my thesis committee, Dr. Mark Scerbo and Dr. Jeremiah Still, for their assistance and guidance for completing my thesis. I would like to thank Linda Forchas, Peggy Kinard, and Mary Boswell, for their help in completing my thesis. Additionally, I would like to thank my lab mates for providing constructive feedback on my thesis presentation and social support during my thesis defense. Finally, and most importantly, I would like to thank my family for their countless support. They are an integral part for completing my thesis.

TABLE OF CONTENTS

Page

LIST OF FIGURES viii

Chapter

I. INTRODUCTION 1
AUTOMATION 2
HUMAN-AUTOMATION TRUST 5
BASIS OF TRUST 7
AUTOMATION TRUST IN MULTITASKING WORKSPACE 8
ATTENTIONAL RESOURCE ALLOCATION 9
CURRENT STUDY 12

II. METHOD 15
PARTICIPANTS 15
APPARATUS 15
FLIGHT SIMULATION TASK 16
TRACKING TASK 17
SYSTEM MONITORING TASK 18
FUEL MANAGEMENT TASK 19
PROCEDURE 20
DEPENDENT VARIABLE 21
SUBJECTIVE WORKLOAD 21
TRUST 21
ATTENTION ALLOCATION 22
MATB PERFORMANCE 22
DESIGN 22
STATISTICAL ANALYSIS 22

III. RESULTS 25
MANIPULATION CHECK 25
EQUAL PRIORITY CONDITION 25
TRACKING PRIORITY CONDITION 26
SUBJECTIVE WORKLOAD 26
CHANCEY ET AL'S (2017) TRUST SCALE 26
PERFORMANCE-BASED TRUST 26
PROCESS-BASED TRUST 27
PURPOSE-BASED TRUST 28
JIAN ET AL.'S (2000) TRUST SCALE 29
ATTENTION ALLOCATION 29
PDT ON TRACKING DISPLAY 31
PDT ON SYSTEM MONITORING DISPLAY 32
PDT ON FUEL MANAGEMENT DISPLAY 33

Page

TRACKING PERFORMANCE 34
SYSTEM MONITORING PERFORMANCE 35
FA EVENTS 35
HIT EVENTS 36
FUEL MANAGEMENT PERFORMANCE 38

IV. DISCUSSION 40
COMPARING WITH PAST RESEARCH 40
INFLUENCE OF TASK PRIORITY 41
PRACTICAL IMPLICATION 45
LIMITATIONS AND FUTURE STUDY 46

V. CONCLUSION 48

REFERENCES 49

APPENDICES 56
A. INFORMED CONSENT DOCUMENT 56
B. DEMOGRAPHICS FORM 58
C. NASA-TASK LOAD INDEX (TLX) WORKLOAD QUESTIONNAIRE 59
D. TRUST QUESTIONNAIRE 60
E. TRUST QUESTIONNAIRE 62

VITA 64

LIST OF FIGURES

Figure Page

1. Levels of automation of decision and action selection 4
2. A model of the experimental setting 16
3. Sample display of MATB-II task 17
4. Descriptive terms for each range of Bayes factor 24
5. Mean performance-based trust ratings 27
6. Mean process-based trust ratings 28
7. Mean purpose-based trust ratings 29
8. Fixation map for tracking priority condition under high task load condition 30
9. Fixation map for tracking priority condition under low task load condition 30
10. Fixation map for equal priority condition under high task load condition 31
11. Fixation map for equal priority condition under low task load condition 31
12. Mean PDT on the central tracking display 32
13. Mean PDT on the system monitoring display 33
14. Mean PDT on the fuel management display 34
15. Mean response time to FA events 36
16. Mean response time to Hit events 37
17. Mean error rate during Hit events 38
18. Mean fuel in Tanks A and Tanks B 39

CHAPTER I

INTRODUCTION

Many professional tasks demand multitasking including both manually performing tasks and monitoring automation such as operating an aircraft (e.g., Billings, 1997), a robotic arm in space teleoperation (e.g., Li et al., 2014) and an air traffic control system (e.g., Loft et al,, 2016). Unfortunately, multitasking environments demand operators to allocate their limited attentional resources effectively to perform the tasks while monitoring automation (e.g., Wickens et al., 2015). Attentional resources are often conceptualized as a limited amount of mental energy that is supplied to support operator's mental activity (Kahneman, 1973; Wickens et al., 2015). According to Wickens et al.'s (2015) general human information-processing (HIP) model, based on the allocation policy and the task demand, attentional resources are allocated to support information-processing stages including sensory processing, perception, response selection, and response execution (Gopher, 1993; Kahneman, 1973; Wickens et al., 2015; Yamani & Horrey, 2018).

According to HIP model, the amount of attentional resources supplied is determined by the mental workload imposed by the operator (Wickens et al., 2015). Mental workload can be defined as "the cost incurred by a human operator to achieve a particular level of performance" (Hart & Staveland, 1988, pp. 140). Mental workload is associated with the task demand while attention is associated with the amount of mental energy supplied. Depending on the amount of attentional resources supplied and the amount of mental workload, task performance can vary systematically (Kahneman, 1973; Wickens et al., 2015). For example, when the available attentional resources exceed the task demand, operators will perform the task successfully. Alternatively, when the task demand exceeds the attentional resource supply, operators will perform the task poorly. Unfortunately, in a multitasking environment, overall task demand can exceed the operator's attentional resource limit, degrading operator performance. Delegating tasks to automated systems is one way to alleviate resource demands in multitasking workspace.

Automation

Automation is defined as a system which performs functions that were previously performed, or those cannot be performed, by human operators (Bainbridge, 1983; Parasuraman et al., 2000). Automation can benefit operators by enhancing task performance (Parasuraman et al., 1996). However, unfortunately, automation can also alter human behaviors in a counterproductive way (Parasuraman & Riley, 1997). For example, according to the National Transportation Safety Board (2013), while approaching the runway at San Francisco International Airport, the pilot operating Asiana Airlines Flight 214 collided with the seawall. The report stated that the pilot descended the aircraft by changing the thrust of the aircraft that gives the forward force of the aircraft, causing the autothrotttle to lose control of the aircraft's airspeed. The pilot failed to recognize the decrease in the airspeed mainly because the pilot misunderstood the function of the autothrotttle, that it would manage the airspeed while landing. Additionally, according to National Transportation Safety Board (2017), a pilot operating Ravn Connect Flight 3153 collided with the terrain in Alaska. The report stated that the pilot disengaged the Terrain Awareness and Warning System (TAWS) resulting in a fatal terrain collision. These two examples highlight that trained pilots can inappropriately understand or use even highly reliable automated systems, contributing to catastrophic accidents.

Parasuraman and colleagues (2000) offer a design guideline for automated systems that support the human information-processing stages at varying levels of automation depending on a given task demand. According to Parasuraman et al. (2000), automation can support the different stages of human information processing with the functions of automation including information acquisition, information analysis, decision selection, and action implementation. The information acquisition function supports the human's sensory processing allowing detection of target stimuli. The information analysis function supports the human's cognitive and perceptual processing by organizing, extrapolating, and integrating input data. The decision and action selection function support the human's decision making by providing augmented or limited decision selection. Finally, the action implementation function supports execution of the selected action for the operator.

Sheridan and Verplank (1978) described the levels of automation specifically at decision and action selection/implementation stage (see Figure 1). At lower levels of the continuum, operator selects a decision and executes the selected decision, and at higher levels of the continuum, the automation executes autonomously in replace of the operators.

HIGH 10. The computer decides everything, acts autonomously, ignoring the human

9. Informs the human only if it, the computer, decides to

8. Informs the human only if asked, or

7. Executes automatically, then necessarily informs the human, and

6. Allows the human a restricted time to veto before automatic execution, or

5. Executes that suggestion if the human approves, or

4. Suggest one alternative

3. Narrows the selection down to a few, or

2. The computer offers a complete set of decison/action alternatives, or

LOW 1. The computer offers no assistance: human must take all decisions and actions

Figure 1. Levels of automation of decision and action selection (Sheridan & Verplank, 1978**)**.

Modern professional environments often involve some levels of automation to enhance overall task performance and efficiency. However, implementing automation has changed the operators' role from actively operating the systems to passively monitoring the system's performance (Bainbridge, 1983), a task that humans perform poorly (e.g., Mackworth, 1948; Molloy & Parasuraman, 1996; Warm, Parasuraman, & Matthews, 2008). To detect system malfunction accurately, *alerted-monitor systems* have been used to present the state of the automation and to direct the operator's attention to system malfunction (Bainbridge, 1983). The alerted-monitor system consists of an automated alerting subsystem which provides automated decisions about the system's state (Sorkin & Woods, 1985). An alerted-monitor system can consist of, for example, an operator performing the system monitoring task and the signaling system detecting engine malfunction that alerts the operator.

When the alerted-monitor system alerts for system malfunction, the operators respond to the alerted-monitor system by analyzing input data and the decisions made by the system. However, due to inherent noise in incoming data, the alerted-monitor system can produce signaling system errors (i.e., false alarms and miss events) depending on the sensor threshold setting (Getty, Swets, Pickett, & Gonthier, 1995). Particularly, a conservative criterion (i.e., when the system requires more sensory evidence to detect system malfunction) for detecting signals would produce higher miss rates while a liberal criterion (i.e., when the system requires less sensory evidence to detect system malfunction) would produce higher false alarm rates. False alarms occur when the alerted-monitor system detects system malfunction given that the system is functioning properly. Miss events occur when the alerted-monitor system fails to detect system malfunction. These signaling system errors could cause the operators to delay their response (Breznitz, 1984; Getty et al., 1995; Sorkin, 1988), increase workload (Dixon & Wickens, 2006), and alter automation use strategies (Lee & See, 2004).

Human-automation Trust

An important factor that can influence human-automation interaction is human-automation trust (Bliss & Dunn, 2000; Chancey et al., 2017; Meyer, 2001; Rice, 2009). Human-automation trust can be defined as “an attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” (Lee & See, 2004, pp. 51). Human-automation trust is critical for successful human-automation interaction due to the increasing complexity and uncertainty induced by the automation (Lee & See, 2004).

Human-automation trust research originated from research on interpersonal trust, and theories of interpersonal trust have been applied to the context of human-automation interaction (Muir, 1994; Muir & Moray, 1996). Muir (1994) developed a theoretical model of trust in automation by integrating Barber's (1983) definition of trust and a model of trust dynamics in interpersonal relations (Rempel et al., 1985). In Barber's (1983) definition, trust develops based on three dimensions including: 1) the expectation that the natural and moral law is constant (i.e., *persistence*), 2) the expectation that the other person possesses the competency for role performance (i.e., *technical competence*), and 3) the expectation that the other person has the moral obligation to fulfill their duty (i.e., *fiduciary responsibility*). In the model of trust dynamics (Rempel et al., 1985), interpersonal trust changes through a series of different dimensions. Initially, interpersonal trust is determined by the extent to which a person's behavior can be anticipated (i.e., *predictability*). As the relationship progresses, interpersonal trust is based on the personal quality of *dependability*, that is the extent to which a person can be counted on. In a matured relationship, interpersonal trust involves the certainty of a person's behavior in the future (i.e., *faith*).

Muir (1994) conceptualized human-automation trust based on the models proposed by Barber (1983) and Rempel et al. (1985). Muir's (1994) model assumes that the dynamics of human-automation trust operate in the same way as the dynamics of interpersonal trust. Thus, when operators have less experience with automation, human-automation trust is mainly controlled by the predictability of the automation's behavior, followed by dependability and then ultimately faith. Also, human-automation trust is based on three different expectations including persistence, technical competence, and fiduciary responsibility. In the context of human-automation interaction, persistence refers to the extent that the natural and moral law stays constant, technical competence refers to the capability that the automation can execute role performance, and fiduciary responsibility refers to the automation's moral obligation to fulfill the operator's duty. Muir's (1994) model predicts the basis of an operator's trust toward automation at varying degrees of experience with automation.

Muir and Moray (1996) tested Muir's (1994) model by conducting two experiments in a supervisory control task. Their task required the participant to maximize the pasteurizer's output by controlling the pump subsystem and the heating subsystem. In both experiments, the pump subsystem was semiautomated allowing the participants to change the pump rate using manual control or automatic control. The results supported Muir's (1994) model that the expectation of technical competence and fiduciary responsibility contributed to human-automation trust. Also, both experiments demonstrated that the three dimensions (i.e., predictability, dependability, and faith) influenced the dynamics of human-automation trust. However, the findings indicated that human-automation trust evolved through the three dimensions in the opposite order from that predicted by Muir's (1994) model. This suggests that the basis of interpersonal trust influenced human-automation trust, but the basis of human-automation trust developed in the opposite order as Rempel et al.'s (1985) model of interpersonal trust (i.e., faith, dependability, and predictability, respectively).

Basis of trust. According to Lee and Moray (1992), human-automation trust is controlled by three separate dimensions including *performance* (i.e., "what the automation is doing"), *process* (i.e., "how the automation is performing"), and *purpose* (i.e., "why the automation was developed"). The performance dimension of trust depends on the capability of the automation to achieve the operator's goal. When the operator believes that the automation has the capability to accomplish the operator's task, performance-based trust develops. The performance-based trust can explain Barber's (1983) competence dimension of trust and Rempel et al.'s (1985) predictability dimension of trust. The process dimension of trust develops when the operator believes that the automation's algorithm is capable of performing a task. The process-based trust can explain Barber's (1983) persistence dimension of trust and Rempel et al.'s (1985) dependability dimension of trust. The purpose dimension of trust depends on the correspondence between the designer's intent and the operator's understanding of the intention. Thus, purpose-based trust develops when the operators accurately understand the designer's intention of developing the automation. The purpose-based trust relates to Barber's (1983) fiduciary responsibility dimensions of trust and Rempel et al.'s (1985) faith dimension of trust.

Automation trust in multitasking workspace. Though ideal automation reduces the operator's mental workload, unfortunately, operators do not necessarily use automation appropriately under high workload conditions. Several studies have examined automation trust in multitasking environment (Bailey & Scerbo, 2007; Karpinsky et al., 2018; Sato et al., 2019). For example, Karpinsky and colleagues (2018) used the MATB paradigm to examine the effects of task load on human-automation trust. Participants were asked to perform a flight simulation task involving the tracking task and the system monitoring task with the assistance of an imperfect signaling system at the reliability of 70%. Workload was manipulated by changing moment-to-moment deviations of the moving circular target from the designated route in the tracking task. Results demonstrated lower levels of trust towards the signaling system under high workload conditions. Further examination of trust was conducted by analyzing the three dimensions of trust proposed by Lee and Moray (1992) and showed that increasing task load led to lower levels of performance- and process-based trust, suggesting that trust towards automation failed to develop potentially due to the misperception of the system's behavior.

Although Karpinsky et al. (2018) demonstrated lower levels of trust toward the automation, there was contrasting evidence on automation trust in multitasking workplace (Bailey & Scerbo, 2007). For example, Bailey and Scerbo (2007) examined the automation trust when performing three different monitoring tasks (i.e., gauge monitoring task, automation mode monitoring task, and digital readout monitoring task) with an assistance of a high (i.e., 98% in Experiment 1 and 99.7% in Experiment 2) or low reliable automation (i.e., 87% in both experiments). Results demonstrated higher levels of trust and poor monitoring performance when operators utilize highly reliable automation for performing monitoring tasks that demand more attentional resources. Although both studies examined different factors that influence human automation trust, the decreased ratings of subjective trust in Karpinsky et al.'s (2018) study could have been accounted by an operator's allocation of attentional resources.

Attentional Resource Allocation

Previous research on human performance in multitasking workspace indicates that operators reduce their attention allocation toward a task monitored by a reliable automated system (Karpinsky et al., 2018; Sato et al., 2019). Importantly, when the operators allocated less attention to monitor the automated system's performance due to increased task demand, they also reported less subjective trust on the performance- and process-dimension of trust, demonstrating that attention is an important factor that influences human-automation trust. It is speculated that the operator's subjective trust decreased due to the operator's failure to switch tasks known as attentional tunneling.

Attentional tunneling is defined as the allocation of attentional resources to a particular task for more than the required time, given that the operator acknowledges the cost of switching to other tasks (Wickens & Alexander, 2009). This phenomenon is presumed to affect multitasking performance. In Karpinsky et al.'s (2018) study, operators took less time to scan the system monitoring display when the central tracking task required more frequent manual input. Thus, operators could have made fewer saccades from the central tracking display indicating a failure to switch to the system monitoring display and vice versa. Several flight-simulated studies have examined the causes of attentional tunneling (Fadden et al., 2001, Fischer et al., 1980; Wickens & Long, 1995). For example, Fischer et al. (1980) examined the pilot's ability to detect unexpected information presented on the outside environment or on the heads-up display (HUD) during an approach. The study examined the deviation of the aircraft from the pathway and the pilot's average response time (RT) for detecting an obstruction on runway. Results demonstrated longer RT for detecting obstruction on the outside environment when using HUD than heads-down display. Furthermore, few pilots were not able to detect obstruction when using HUD. Fischer et al. (1980) reasoned that the symbol on the HUD inhibits the pilots from looking outside the environment since the symbols make immediate changes that can be easily perceived by the pilot. Similar studies examined the attentional tunneling by employing a low-fidelity (Wickens & Long, 1995) and high-fidelity flight simulator (Fadden et al., 2001).

In applied environments, eye movements are often used as a measure of overt attention. Though the location of a fixation and that of attention can be decoupled (e.g., Posner et al., 1980), monitoring behavior of a set of dynamic areas of interest is well explained by models of supervisory control that assumes the coupling of focal vision and attention. More specifically, due to the restricted range of focal vision within the fovea strongly associated with attention, the eye can be considered a single-server queue. That is, objects in a visual scene await to be serviced in a queue, and eye movements are considered a way to service this queue (e.g., Moray, 1986; Senders, 1964). In the current context, percent dwell time in each areas of interest (AOIs) is used as a measure of attentional resources (e.g., Salvucci & Taatgen, 2008). For example, Karpinsky and colleagues (2018) examined the participant's eye movements in the MATB paradigm involving the central tracking task and the system monitoring task. Their results demonstrated that the operators attended less to the system monitoring task assisted by an automated system when the tracking task demanded more frequent input.

Theoretical and computational modeling approaches lend support for the control of attention allocation in dynamic multitasking visual environments. The Strategic Task Overload Management (STOM) model (Wickens et al., 2013) predicts the operator's likelihood of switching away from the ongoing task in a multitasking environment with partial assistance of the automation. The model involves top-down and bottom-up factors that influence the participant's attention allocation which offers theoretical frameworks for the current study.

More specifically, the STOM (Wickens et al., 2013) describes the operator's task switching behavior of attending from an ongoing task (OT) to an alternative task (AT) in a multitasking environment. According to the model, attention switches from an OT to an AT when an AT has higher "attractiveness." The attractiveness of an AT is determined by the extent to which a task can direct an operator's attention to itself (i.e., salience), the value of a task (i.e., priority), the extent to which a task is engaged (i.e., task interest), and the amount of effort required by the task (i.e., task difficulty; Wickens et al., 2013). In Wickens and colleagues' (2016) study, priority was shown to have minimal influence on the attractiveness of the task. Thus, the model was revised as follows:

$$\text{Attractiveness} = I + S + D$$

where I refers to task interest, S denotes salience, and D represents task difficulty. Once the attention switches to the AT, the AT then becomes the OT. A validation study (Wickens et al.,

2016) of the new model provided better prediction for task attractiveness. The model accounted for 95% of variance when priority was excluded from the model.

Current Study

It remains unclear why the central tracking demand reduced their attention towards the imperfect automation and automation trust in Karpinsky et al.'s (2018) study. One possibility is that the participants might have increased their task priority to the central tracking task in the high load condition where the tracking task required frequent and repetitive manual correction of the path of the aircraft. Further, attentional tunneling impaired their perception of behaviors of the signaling system. Although Wickens et al. (2016) demonstrated minimal influence of task priority, a large effect was observed in previous research (Gopher et al., 1982). Manipulation of task priority in Wickens et al. (2016) might have failed potentially due to the absence of the baseline performance. In Gopher et al.'s (1982) study, for example, participants received continuous feedback on their tracking performance during the practice session reflecting their baseline performance. As a result, participants were able to prioritize the tracking task at a level of 30%, 50%, and 70%. Conversely, Wickens et al. (2016) verbally instructed the participants to prioritize the tracking task or equally prioritize all task without an anchor reflecting their baseline performance. Thus, it is expected that following Gopher et al.'s (1982) procedure could allow successful manipulation of task priority in the current experiment.

The present study aimed to replicate Karpinsky et al. (2018) and examined the effect of task priority on attention allocation on automation trust towards imperfect signaling system with a reliability of 70% in the MATB paradigm. Participants were asked to perform three concurrent tasks, the fuel management task, the system monitoring task, and the central tracking task. The system monitoring task was assisted by the automated signaling system. For the central tracking task, participants performed the task with low and high difficulty. Participants were instructed to prioritize either the central tracking task over the other two tasks (tracking priority condition) or maximize performance for all tasks (equal priority condition).

It was hypothesized that the present study would demonstrate the main effect of task load based on Karpinsky et al.'s (2018). More specifically, I hypothesized that:

1. Participants would fixate less frequently towards the system monitoring task in the high tracking difficulty condition (i.e., high task load) than the low tracking difficulty condition (i.e., low task load).
2. Participants would report higher levels of workload in the high tracking difficulty condition than the low tracking difficulty condition.
3. Participants would make more errors in the system monitoring task and show longer RT towards the automated signaling system in the high tracking difficulty condition than the low tracking difficulty condition.
4. Participants would present poor control of the circular moving target (i.e., poor tracking performance) in the high tracking difficulty condition than the low tracking difficulty condition.
5. Participants would report lower levels of trust towards the automated signaling system in the high tracking difficulty condition than the low tracking difficulty condition, on the performance and process dimensions but not on the purpose dimension.

Based on previous work (Gopher et al., 1982), the current study would yield the main effect of task priority. Specifically, I hypothesized that:

6. Participants would fixate less frequently towards the system monitoring task when the central tracking task was prioritized over all other tasks (i.e., tracking priority condition) than when all tasks are equally prioritized (i.e., equal priority condition).
7. Participants would make more errors in the system monitoring task and show longer RT towards the automated signaling system when central tracking task was prioritized over all tasks than when all tasks are equally prioritized.
8. Participants would present poor control of the circular moving target when all tasks are equally prioritized than when the central tracking task was prioritized over all other tasks.

Finally, if task priority is a factor that influenced attention allocation and automation trust in Karpinsky et al. (2018), then the equal priority instruction should eliminate the effect of task load on attention allocation and automation trust. Specifically, I hypothesized that:

9. The effect of the tracking task on the four variables in Hypotheses 1-5 would be present in the tracking priority condition but eliminated in the equal priority condition.

CHAPTER II

METHODS

Participants

Based on Karpinsky et al. (2018), 40 participants (27 females, mean age = 21.54 years, SD = 8.62) were recruited from Old Dominion University (ODU). All participants were screened for normal or corrected-to-normal vision and normal color perception using the Ishihara Color Blindness test (Ishihara, 2014). Participants received course credits via the ODU SONA system for their participation.

Apparatus

Stimuli were presented on a Samsung T24C550 23.6" LED monitor (1920 x 1080) with a frame rate of 75 Hz. Windows 7 (Dell OptiPlex 9020) was used to run the MATB-II (Santiago-Espada, Myer, Latorella, & Comstock, 2011) which is a computer-based flight simulation program that assesses flight performance. Participant's eye movements were recorded using the Eyelink II (SR Research, Mississauga, Ontario, Canada) with a sampling rate of 250 Hz. Participants fixed their head on the chin rest placed approximately 80 cm away from the monitor. The experiment was conducted in a quiet room with dim light. Figure 2 present a model of the experimental setting.

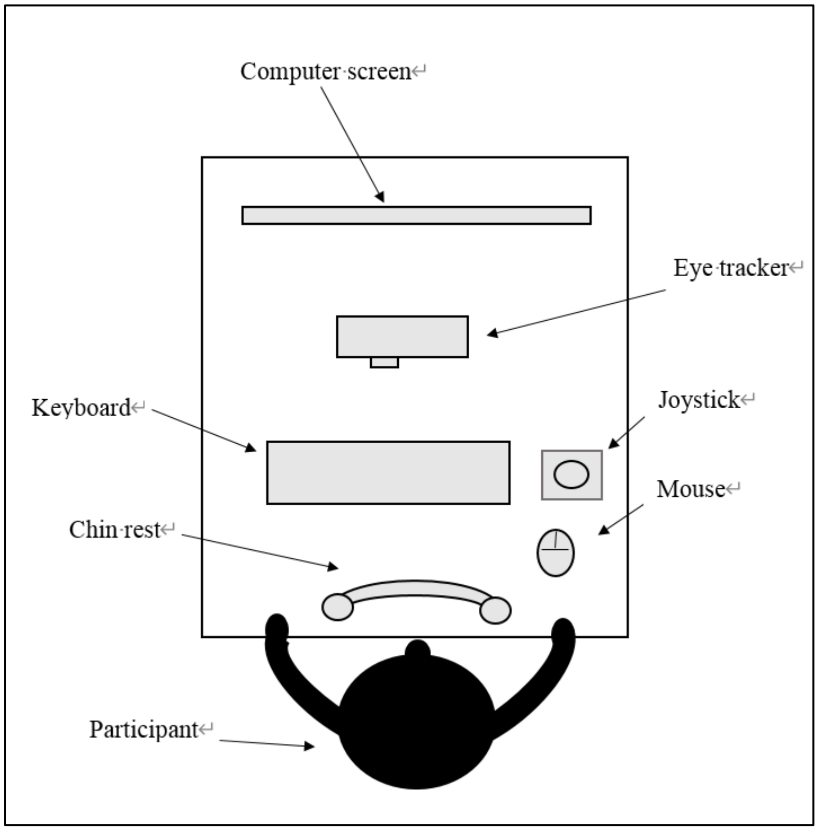

Figure 2. A model of the experimental setting.

Flight Simulation Task

The current experiment involves the use of an automated signaling system in a flight simulation task, known as the Multi-Attribute Task Battery-II (MATB-II; Santiago-Espada et al., 2011). In the MATB-II, the signaling system incorporates information acquisition, information analysis, and decision selection supports but not action implementation. Operators are asked to monitor four gauges that represent performance of four engines of a simulated aircraft and the automated signaling system alerts when the engine malfunctions. The signaling system uses the information acquisition function to help operators detect the malfunction by acquiring input data from the aircraft's engine. After receiving input data, the signaling system uses the information analysis function to organize input data to project the state of the aircraft's engine. Then, the signaling system uses the decision and action selection function to recommend a decision for the operators to execute, but the signaling system does not have an action implementation function because operators must manually acknowledge and correct gauges following each alert. The MATB-II consisted of four different tasks including the central tracking task, the system monitoring task, communication task, and the fuel management task. Participants in the equal and tracking priority condition performed the central tracking task, the system monitoring task, and the fuel management task. Figure 3 presents the flight-simulation task.

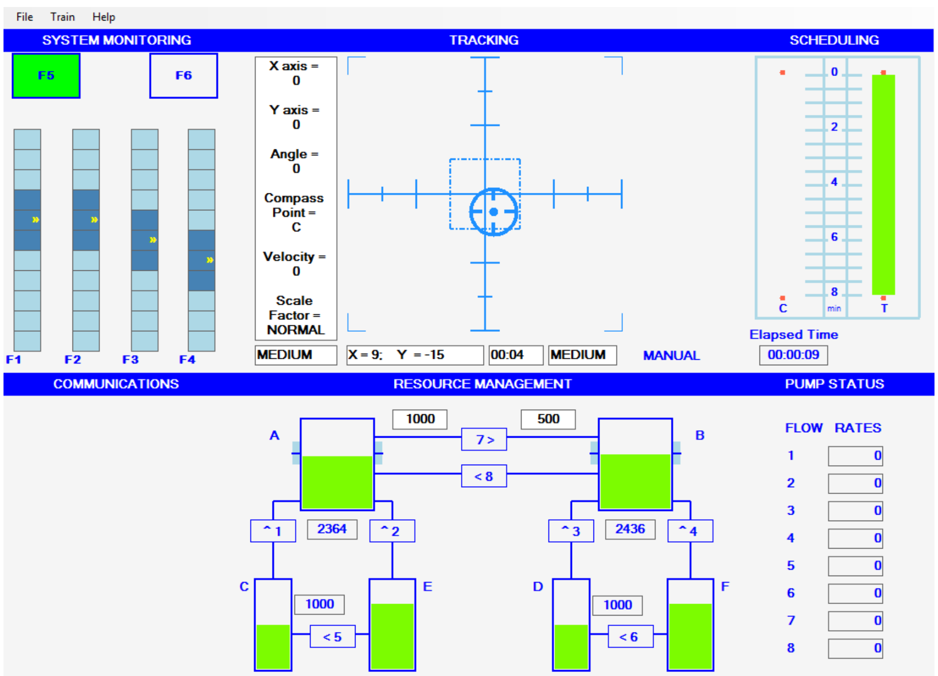

Figure 3. Sample display of MATB-II task. System monitoring (top left), central tracking task (top center), and fuel management task (bottom center).

Tracking task. Participants were required to control the moving circular target within the dotted square using a joystick whenever the moving circular target deviated from the tracking display (top center of Figure 5). The moving circular target represented the direction of the aircraft while the dotted square represented the designated route. The deviation of the moving circular target from the dotted square was controlled by setting the frequency of the force function. The frequency of the force function between the two experimental trials differed since the difficulty of the central tracking task was manipulated as a within-subjects factor. For the high task load condition, the frequency was set to 0.12 Hz while the frequency for the low task load condition was set to 0.06 Hz. However, in the practice session, the frequency for the central tracking task was set to 0.09 Hz. The root mean squared (RMSE) deviation between the moving circular target and the dotted square was calculated by sampling the input in X and Y dimensions at 20 Hz. For each block, the mean of RMSE was calculated to assess the performance for the central tracking task.

System monitoring task. In the system monitoring task (top left of Figure 5), participants were asked to monitor the four vertical gauges, which represented the state of the aircraft's four engines, and corrected the vertical fluctuating pointer when the pointer hit either the upper or lower extremity. The alarm box, located above the gauges, served to notify participants by presenting two different states of the engines. The green box indicated that the engine is in "normal" state. The green box illuminated until one of the pointers within each gauge hit the extremity of the gauge. The red box indicated that one of the engines is in "warning" state. The red box illuminated when one of the pointers within each gauge hit either extremities of the gauge. Each block consisted of 28 hit (Hit) events and 12 false alarm (FA) events which occurred in random intervals so that the reliability was 70%. Miss events were excluded from this study since performance and trust did not differ between FA events and Miss events (see Karpinsky et al., 2018). Hit events occurred when the signaling system correctly detected engine malfunction while FA events occurred when the signaling system detected engine malfunction even though the engine was functioning normally. In a Hit event, one of the fluctuating pointers hit the upper or lower extremity of the gauge which turned the green box off and turned the red box on. This indicated a change in the engine's state from a "normal" state to a "warning" state. In this case, participants were instructed to respond to the signaling system by using the mouse to click the red box and the green box. Also, participants were required to reset the fluctuating pointer by using the mouse to click the corresponding gauge which was labeled F1 to F4. In a FA event, the green box was turned off and the red box was turned on even though the fluctuating pointers did not hit the upper or lower extremity of the gauge. In this case, the participants only responded to the signaling system. Participants were not required to reset the fluctuating pointer since none of the pointers hit the extremity of the gauge. Participant's data were excluded from the analysis when the error rate is above 50%.

Fuel management task. The fuel management task (bottom center of Figure 5) required participants to maintain fuel in Tanks A and Tanks B, located adjacent to letters A and B, respectively. In each block, the fuel in Tanks A depleted at a rate of 1,000 units per minute while Tanks B depleted at a rate of 500 units per minute. When one of the tanks consumed below a volume of 2,500 units, participants were instructed to transfer fuel from lower supply tanks, located adjacent to letters C to F, by activating one of the pumps, labeled with numbers 1 to 8. An arrow was placed next to each number indicating the direction of the pump flow. The pumps could be activated by using the mouse to click the number of the corresponding pumps on the computer screen. The pump flow rate for all pumps was set to 900 units per minute. The task included a pump failure event in which one of the pumps failed to activate for 10s. In this case, participants were required to compensate by activating other pumps. Each block consisted of 8 pump failure events which occurred in random intervals. The status of the pump was denoted by the pump's color. A green numbered box indicated that the pump was activating. A white numbered box indicated that the pump was deactivated. A red numbered box indicated that the pump was experiencing pump failure.

Procedures

Participants completed an informed consent and a demographics form, followed by screening for color perception and visual acuity. Then, participants were randomly assigned to either the equal or tracking priority condition. Following the procedure of Gopher et al. (1982), participants in the tracking priority condition received an instruction asking them to prioritize the central tracking task at a priority level of 70%, meaning that they were encouraged to perform a level better than the lowest 70% of their own baseline level performance. On the other hand, participants in the equal priority condition received an instruction asking them to prioritize the central tracking task at a priority level of 30%. Participants in both conditions were instructed to perform the flight-simulation task with a single hand. In the practice session, participants performed each flight-simulation task separately for a total of nine minutes (i.e., three minutes for each task) and all flight-simulation task for 3 minutes simultaneously. After the practice session, participants received a value representing their baseline performance distribution based on their own performance during the practice session for the central tracking task and a target value for the experimental session through a computer screen. The participant's baseline performance distribution was determined by calculating the average RMSE. The target value for participants in the equal priority condition was calculated by adding the average RMSE by one standard deviation. On the other hand, the target value for participants in tracking priority condition was calculated by subtracting the average RMSE by on standard deviation.

Prior to the experimental session, the eye tracker was calibrated using the standard 9-dot calibration system. In the experimental session, participants completed two 20-minute experimental blocks which were counterbalanced to reduce the impact of potential order effects. After each block, participants completed Chancey et al.'s (2017) trust questionnaire, Jian et al.'s (2000) trust questionnaire, and the NASA-TLX (Hart & Staveland, 1988). Credits were given for participation. The present study was approved by Institutional Review Board (IRB) at ODU.

Dependent Variables

Subjective workload. The NASA-TLX (Hart & Staveland, 1988) was used to measure subjective workload. Participant's subjective workload of the task was measured based on the ratings of each subscale of workload including mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX consists of 6 items on a 21-point gradient scale ranging from *very low* to *very high.* Thus, the overall minimum score that a participant can rate is 6 while the overall maximum score that a participant can rate is 126. Previous studies have demonstrated test/retest reliability (Hart & Staveland, 1988), convergent validity (Rubio et al., 2004), and concurrent validity (Rubio et al., 2004).

Trust. Chancey et al.'s (2017) and Jian et al's (2000) trust questionnaires were used after each experimental trial. Chancey et al.'s (2017) questionnaire comprised of 13 items on a 12-point Likert scale ranging from (1) *not descriptive* to (12) *very descriptive.* Thus, the overall minimum score that a participant can score is 13 while the overall maximum score is 156. The items are divided into three subsets (i.e., performance, process, and purpose). Note that Chancey et al.'s (2017) trust questionnaire was developed based on Madsen and Gregor's (2000) Human-Computer Trust Questionnaire based on the three-dimensional theory of human-automation trust (Lee & Moray, 1992). Chancey et al.'s (2017) demonstrated high Cronbach's alpha for overall trust rating (α = .97), performance-based rating (α = .96), process-based rating (α = .91), and purpose-based rating (α = .93) indicating internal consistency. In a recent study, a multi-level confirmatory factor analysis of Chancey et al.'s (2017) questionnaire has shown to have a three-dimension structure (Yamani et al., in preparation).

Jian et al's (2000) trust questionnaire is comprised of 12 items on a 7-point Likert scale ranging from (1) *not at all* to (7) *extremely.* The minimum score is 12 while the maximum score is 84. Safar and Turner (2005) has shown a high Cronbach's alpha in Jian et al.'s (2000) questionnaire (α = .93)

Attention allocation. Percentage dwell time (PDT) for each AOI were calculated by examining the proportion of time that the participants fixated on a particular AOI.

MATB-II performance. Tracking performance was examined by calculating the mean of RMSE for each block. System monitoring performance was examined by calculating the error rates and reaction times (RTs) for each experimental trial for both Hit and FA events. Error rates are defined as the proportion of events that the participant responded incorrectly. RTs are defined as the time it takes for the participants to respond correctly after the onset of an event.

Design

The present study employed a 2 x 2 mixed design with task priority as a between-subjects factor and task load as a within-subjects factor. The dependent variables were subjective workload, subjective trust, attention allocation, tracking performance, system monitoring performance, and fuel management performance.

Statistical Analysis

The present study used a 2 x 2 mixed Bayesian analysis followed by Bayesian t-tests. The measures of evidence for Bayesian analysis is Bayes factors, the likelihood ratios reporting the degree to which obtained data favor one of two statistical models against the other. Bayesian analysis circumvents some critical issues with the null-hypotheses significance tests (NHSTs). Unlike p-values in the NHSTs, Bayes factors can provide evidence for or against the effect of interest (Jeffreys, 1961). For instance, a Bayes factor favoring a model without an effect of interest over a model with the effect indicates evidence against the presence of the effect. However, non-significance in the NHSTs does not indicate the absence of an effect of interest. Bayes factor provides a measure of the strength of an effect of interest (Wetzel et al., 2011). For example, a Bayes factor of 3, which is taken as a substantial evidence, indicates that it is 3 times more likely that the supported statistical model generated the observed effect than the unsupported model. The default Bayesian tests (Rouder & Morey, 2012) were used in the current study. Bayes factor values greater than 1 indicate evidence in favor of an intervention effect and against the null, while values less than 1 indicate evidence in favor of the null and against the intervention effect. Following Rouder and Morey (2012), B_{10} denotes these values. Lastly, the Bayes factors were interpreted against the descriptive terms suggested by Jeffrey (1961; Figure 4).

B_{10}	Interpretation
>100	Decisive evidence for H1
30-100	Very strong evidence for H1
10-30	Strong evidence for H1
3-10	Substantial evidence for H1
1-3	Not worth more than a bare mention
1/3-1	Not worth more than a bare mention
1/10-1/3	Substantial evidence for H0
1/30-1/10	Strong evidence for H0
1/100-1/30	Very strong evidence for H0
<1/100	Decisive evidence for H0

Figure 4. Descriptive terms for each range of Bayes factor (Jeffrey, 1961).

CHAPTER III

RESULTS

The present study employed a 2 x 2 mixed Bayesian analysis of variance (ANOVA) with Task Load (High vs. Low) as a within-subject factor and Task Priority (Equal vs. Tracking) as a between-subject factor. As a manipulation check, a Bayesian paired samples t-test was employed to compare the participant's tracking performance with the participant-specific target value for the central tracking task for each block. Prior to the planned analyses, a 2 x 2 mixed Bayesian ANOVA including Order as an additional between-subject factor to explore the order effect, and the results indicate no evidence for the presence of the order effect for any of the dependent variables.

Six participants were excluded from the current analysis. Of these, two were excluded because the system monitoring performance was below the inclusion criteria (i.e., 50%), three were excluded because of technical issues with the eye tracker, and one withdrew from the study because the participant felt sick during the experimental block.

Manipulation Check

Equal priority condition. Data gave no evidence that the participants in the equal priority condition performed above the target value under high task load condition, suggesting that participants were able to perform at the priority level of 30% in high task load condition [$t(16) = -1.88$, $B_{10} = 1.04$, $d = 0.55$]. However, data gave decisive evidence that the participants performed below the target value under low task load condition, suggesting that participants performed at a priority level more than 30% in low task load condition [$t(16) = 6.78$, $B_{10} = 4.60 \times 10^3$, $d = 1.77$].

Tracking priority condition. Data gave decisive evidence that the participants in the tracking priority condition performed above the target value under the high task load condition, suggesting that participants performed at the priority level less than 70% in high task load condition [$t(16) = -7.90$ $B_{10} = 2.62 \times 10^4$, $d = 2.04$]. Data gave substantial evidence that the participants in the tracking priority condition performed below the target value under the low task load condition, suggesting that participants performed at the priority level more than 70% in low task load condition [$t(16) = 3.18$ $B_{10} = 8.43$, $d = 0.66$].

Subjective Workload

Data indicated strong evidence for the main effect of task load on subjective workload, suggesting a successful manipulation of tracking difficulty [$M = 77.00$ vs. 66.00 for the high task load condition and the low task load condition, respectively; $F(1, 32) = 20.70$, $B_{10} = 376.00$, $\eta^2_G = 0.10$]. However, data gave no substantial evidence for the main effect of task priority [$F < 1$, $B_{10} = 1/2.80$] and the interaction effect [$F < 1$, $B_{10} = 1/2.19$].

Chancey et al.'s (2017) Trust Scale

The three dimensions of trust in Chancey et al.'s (2017) questionnaire were examined separately. Figures 5, 6, and 7 present mean ratings for performance-, process-, and purpose-based trust, respectively.

Performance-based trust. Participants reported substantially lower levels of performance-based trust in the high task load condition than the low task load condition [$M = 44.18$ vs. 41.62; $F(1, 32) = 6.50$, $B_{10} = 3.24$, $\eta^2_G = 0.01$], replicating the result of Karpinsky et al. (2018). Additionally, task priority substantially modulated the main effect of task load [$F(1, 32) = 6.50$, $B_{10} = 3.46$, $\eta^2_G = 0.01$]. Follow up t-tests indicated strong evidence that participants rated higher performance-based trust under low task load conditions when all tasks were equally prioritized [$M = 45.53$ vs. 40.41; $t(16) = -3.56$, $B_{10} = 16.64$, $d = 0.39$]. However, data gave substantial evidence that performance-based trust rated under high and low task load conditions did not vary in the tracking priority condition [$t(16) = 0$, $B_{10} = 1/4.00$]. The main effect of task priority was not reliable [$F < 1$, $B_{10} = 1/1.67$].

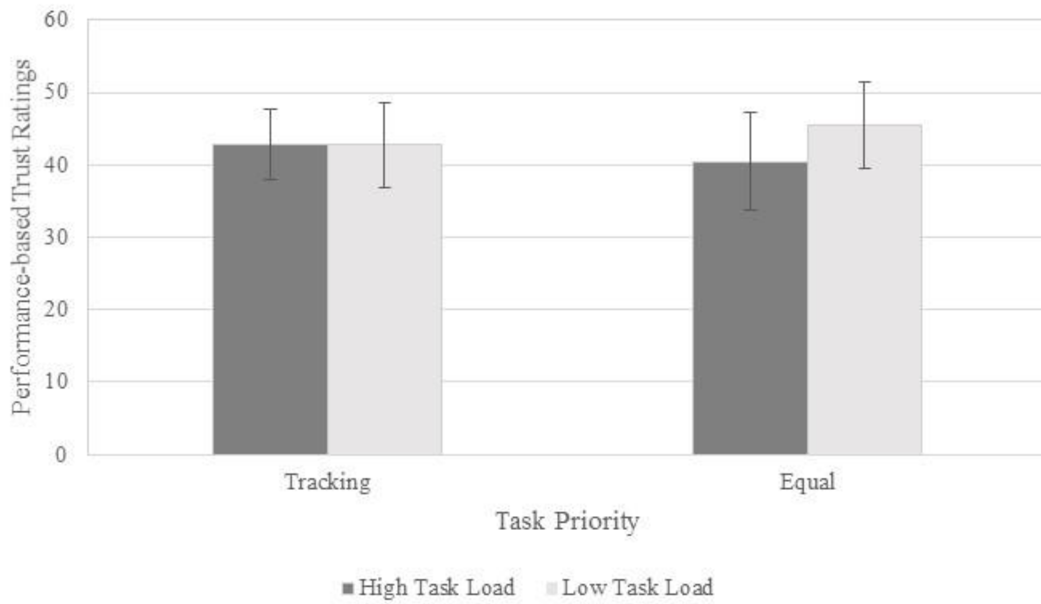

Figure 5. Mean performance-based trust ratings. Error bars represent 95% between-participant confidence intervals.

Process-based trust. Data strongly showed that the high task load condition reduced process-based trust compared to the low task load condition [$M = 44.18$ vs 47.97; $F(1, 32) = 11.58$, $B_{10} = 22.86$, $\eta^2_G = 0.03$]. However, data did not give substantial evidence for the main effect of task priority [$F < 1$, $B_{10} = 1/1.40$] and the interaction effect [$F < 1$, $B_{10} = 1/2.61$].

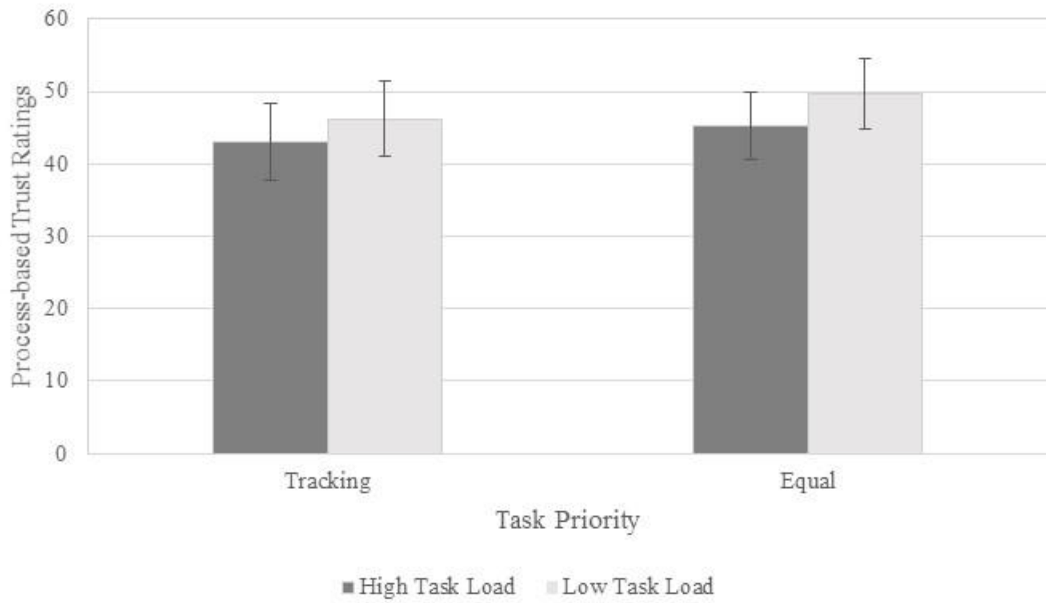

Figure 6. Mean process-based trust ratings. Error bars represent 95% between-participant confidence intervals.

Purpose-based trust. Data indicate no substantial evidence for the main effect of task load [M = 26.44 vs. 25.09 for the low task load condition and the high task load condition, respectively; $F(1, 32) = 3.92$, $B_{10} = 1.22$, $\eta^2_G = 0.01$], the main effect of task priority [$F < 1$, $B_{10} = 1/1.59$, $\eta^2_G = 0.02$], and the interaction effect [$F(1, 32) = 4.27$, $B_{10} = 1.62$, $\eta^2_G = 0.01$].

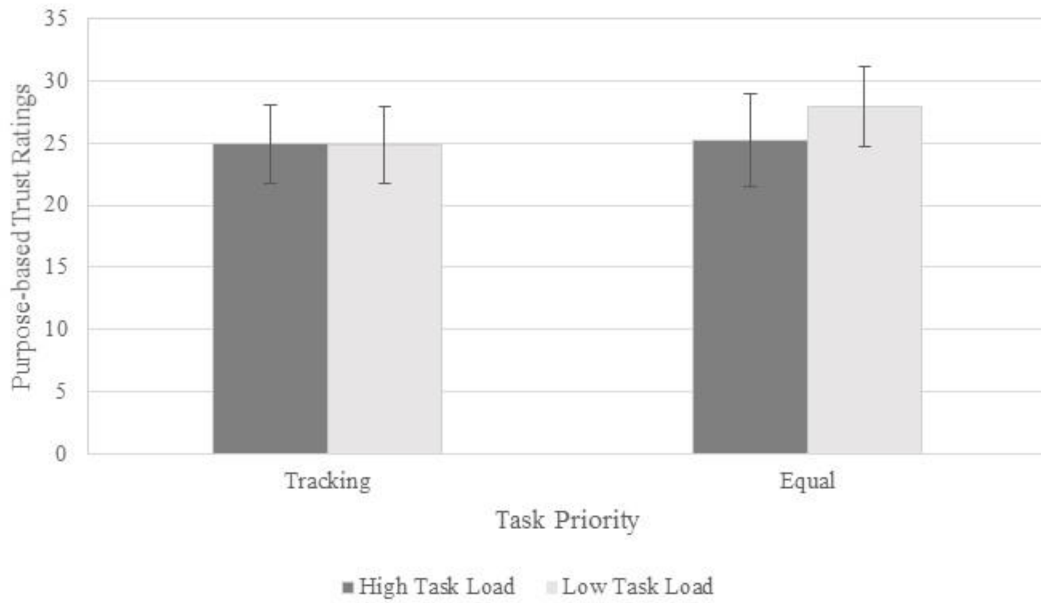

Figure 7. Mean purpose-based trust ratings. Error bars represent 95% between-participant confidence intervals.

Jian et al.'s (2000) Trust Scale

Data on Jian et al.'s (2000) trust questionnaire did not indicate substantial evidence for the main effect of task load [$F(1, 32) = 3.05$, $B_{10} = 1/1.10$, $\eta^2_G = 0.01$], the main effect of task priority [$F < 1$, $B_{10} = 1/2.02$], and the interaction effect [$F(1, 32) = 1.13$, $B_{10} = 1/1.95$, $\eta^2_G < 1$].

Attention Allocation

PDT on all three MATB displays were examined separately. Figures 8 and 9 present fixation maps of a representative participant in tracking priority condition for the high and low task load, respectively. Figures 10 and 11 present fixation maps of a representative participant in equal priority condition for the high and low task load, respectively. The fixation map collectively represents the participant's fixation. Within the MATB display, the red region indicates frequently fixated region while the green region indicates less frequently fixated region.

Figures 12, 13 and 14 present PDT on the central tracking display, the system monitoring display, and the fuel management display, respectively.

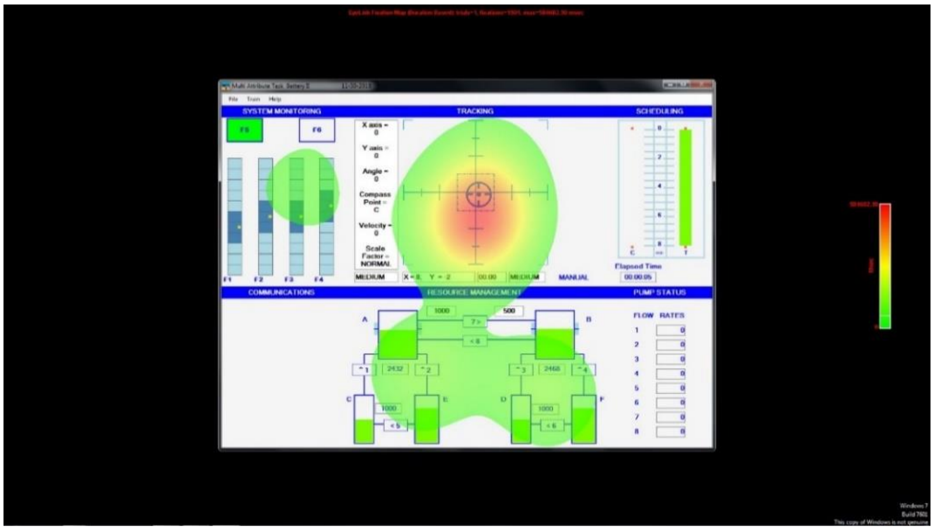

Figure 8. Fixation map of a representative participant in tracking priority condition under high task load condition.

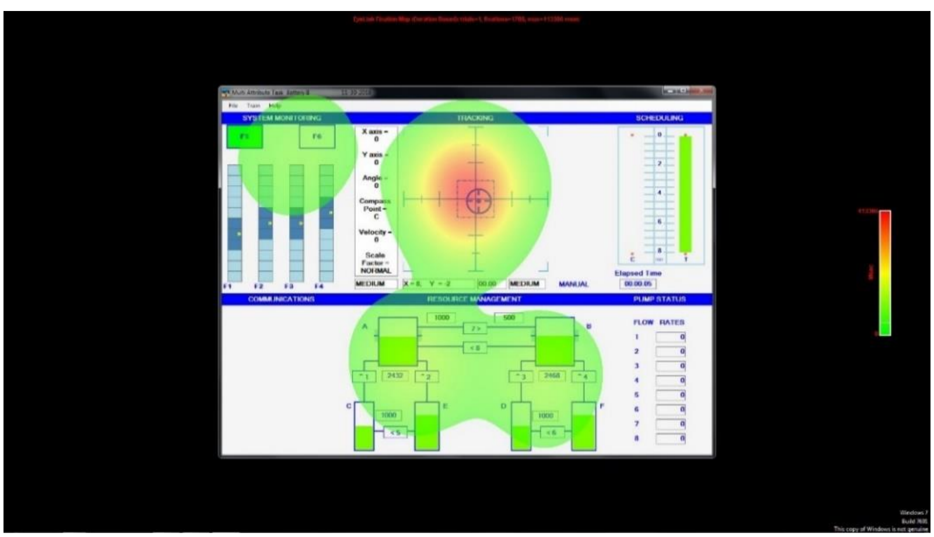

Figure 9. Fixation map of a representative participant in tracking priority condition under low task load condition.

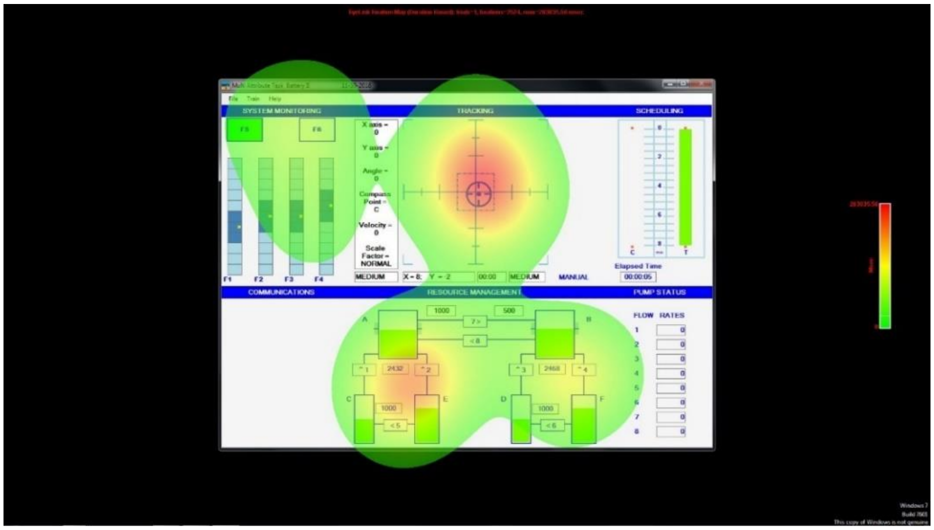

Figure 10. Fixation map of a representative participant in equal priority condition under high task load condition.

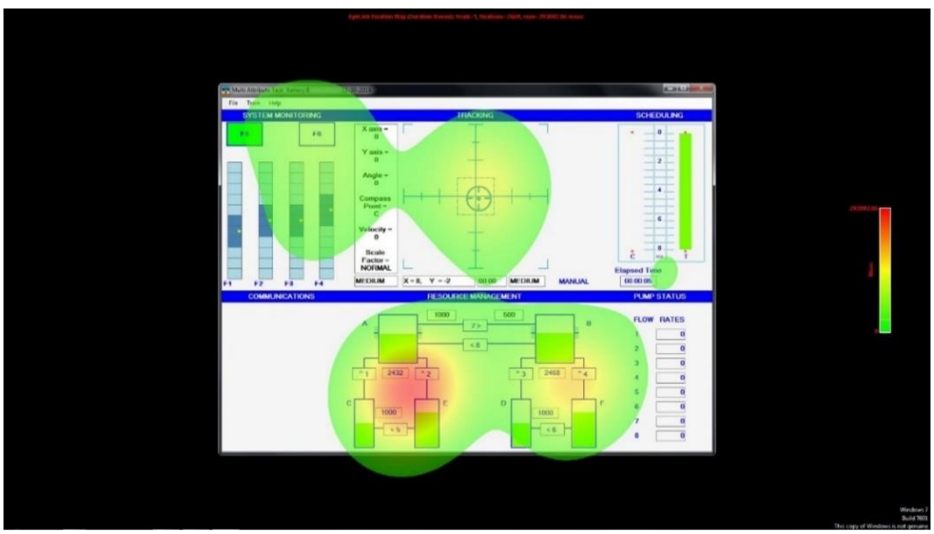

Figure 11. Fixation map of a representative participant in equal priority condition under low task load condition.

PDT on tracking display. Data gave decisive evidence that participants fixated more frequently on the tracking display when the central tracking task demanded more manual corrections [M = 0.46 vs. 0.32 for the high task load condition and the low task load condition, respectively; $F(1, 32) = 45.16$, $B_{10} = 6.85 \times 10^4$, $\eta^2_G = 0.16$]. Furthermore, participants fixated decisively more on the tracking display when the central tracking task was prioritized than underprioritized [M = 0.51 vs. 0.27; $F(1, 32) = 22.85$, $B_{10} = 466.52$, $\eta^2_G = 0.38$]. There was no evidence for the presence or absence of the interaction effect [$F(1, 32) = 1.93$, $B_{10} = 1/1.42$, $\eta^2_G = 0.01$].

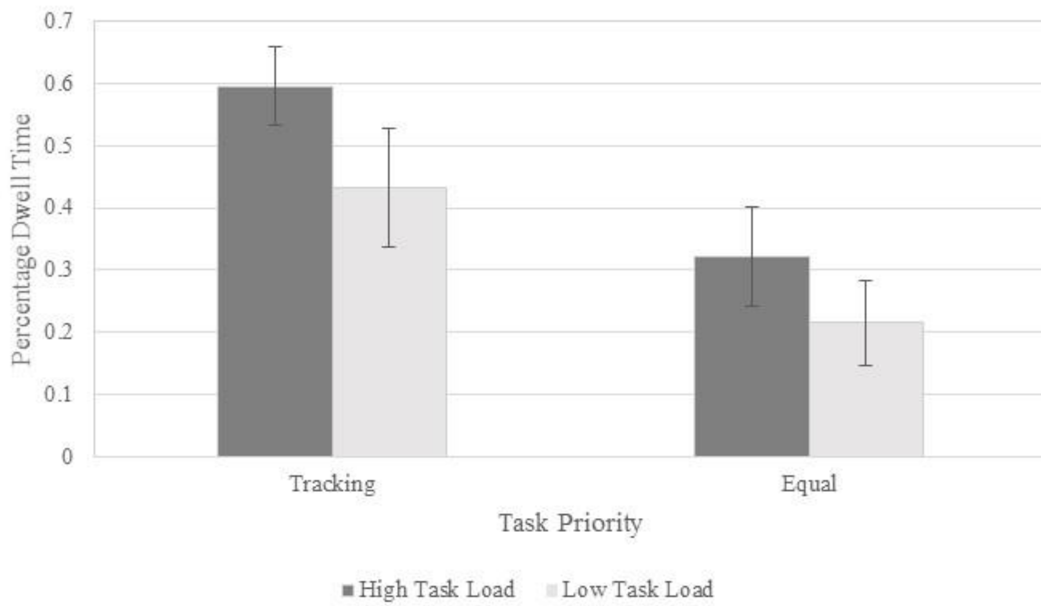

Figure 12. Mean PDT on the central tracking display. Error bars represent 95% between-participant confidence intervals.

PDT on system monitoring display. Data strongly evinced that the high task load condition reduced fixations within the system monitoring display compared to the low task load condition [M = 0.13 vs. 0.11; $F(1, 32) = 11.05$, $B_{10} = 15.22$, $\eta^2_G = 0.04$]. Additionally, participants fixated substantially more on the system monitoring display when the central tracking task was less prioritized [$M = 0.14$ vs. 0.10 for equal priority condition and tracking priority condition, respectively; $F(1, 32) = 7.94$, $B_{10} = 5.45$, $\eta^2_G = 0.18$]. Again, no interaction effect was substantial [$F < 1$, $B_{10} = 1/2.96$].

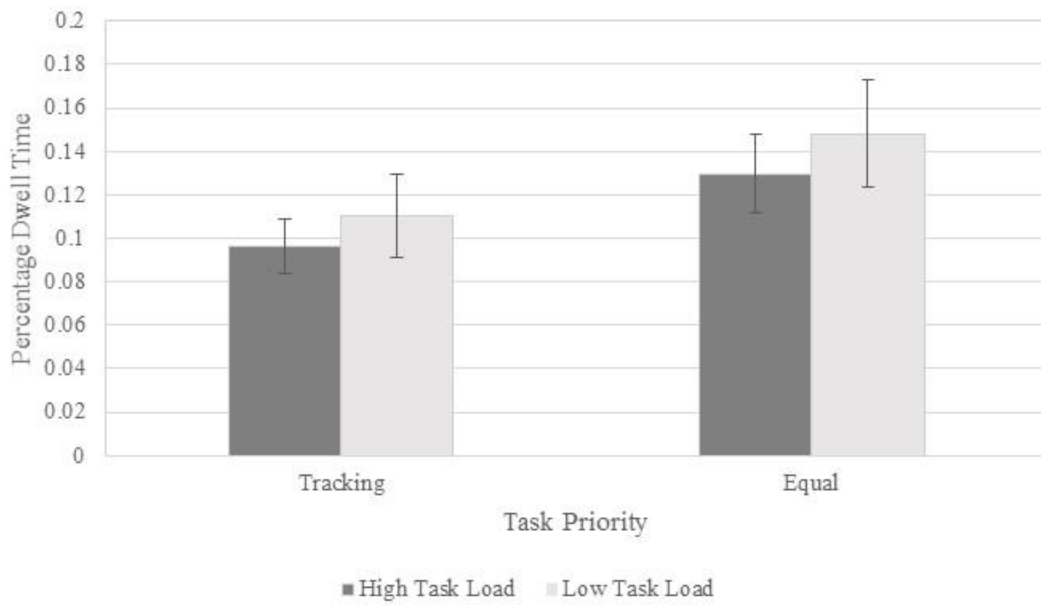

Figure 13. Mean PDT on the system monitoring display. Error bars represent 95% between-participant confidence intervals.

PDT on fuel management display. Data indicated decisive evidence that participants fixated more on the fuel management display when the central tracking task required more corrective input [$M = 0.45$ vs. 0.36 for the high task load condition and low task load condition, respectively; $F(1, 32) = 34.52$, $B_{10} = 7.30 \times 10^3$, $\eta^2_G = 0.09$]. Data showed very strong evidence that participants fixated less on the fuel management display when the central tracking task was more prioritized [$M = 0.31$ vs. 0.50 for the tracking priority condition and equal priority condition respectively; $F(1, 32) = 15.89$, $B_{10} = 53.37$, $\eta^2_G = 0.31$]. The interaction effect was not substantial [$F < 1$, $B_{10} = 1/2.43$].

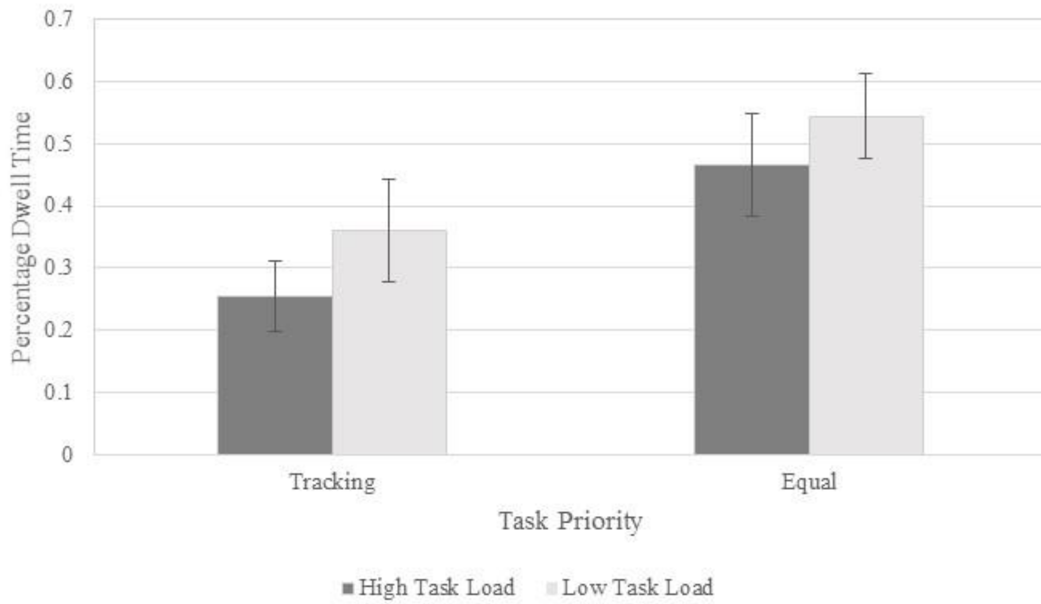

Figure 14. Mean PDT on the fuel management display. Error bars represent 95% between-participant confidence intervals.

Tracking Performance

The participants' tracking performance measured on RMSE was decisively lower in the high task load condition than the low task load condition [$M = 45.14$ vs. 24.99; $F(1, 32) = 230.11$, $B_{10} = 2.90 \times 10^{14}$, $\eta^2_G = 0.53$], suggesting greater deviation of the cursor from the target when the central tracking task required more corrective input. Furthermore, tracking performance declined when the tracking task was less prioritized [$M = 41.28$ vs. 28.84 for the equal priority condition and tracking priority condition, respectively; $F(1, 32) = 16.16$, $B_{10} =$

71.06, $\eta^2_G = 0.30$]. Data gave no substantial evidence for or against the interaction effect [$F(1, 32) = 2.88$, $B_{10} = 1/1.12$, $\eta^2_G = 0.01$].

System Monitoring Performance.

FA events. Figure 15 presents mean response time to FA events. For the response times, data provided very strong evidence that participants responded faster to FA events in the low task load condition than the high task load condition [$M = 3.09$ vs. 3.62 seconds; $F(1, 32) = 16.23$, $B_{10} = 63.52$, $\eta^2_G = 0.06$] and when lower priority was placed in the tracking task [$M = 2.48$ vs. 4.24 seconds for the equal priority and the tracking priority condition, respectively; $F(1, 32) = 23.02$, $B_{10} = 430.06$, $\eta^2_G = 0.39$]. There was anecdotal evidence for the interaction effect [$F(1, 32) = 5.46$, $B_{10} = 2.18$, $\eta^2_G = 0.02$].

For the error rates, data gave substantial evidence against the main effect of task load [$F < 1$, $B_{10} = 1/3.37$], and only anecdotal evidence against the main effect of task priority [$F(1, 32) = 2.96$, $B_{10} = 1/1.05$, $\eta^2_G = 0.05$], and the interaction effect [$F < 1$, $B_{10} = 1/2.94$].

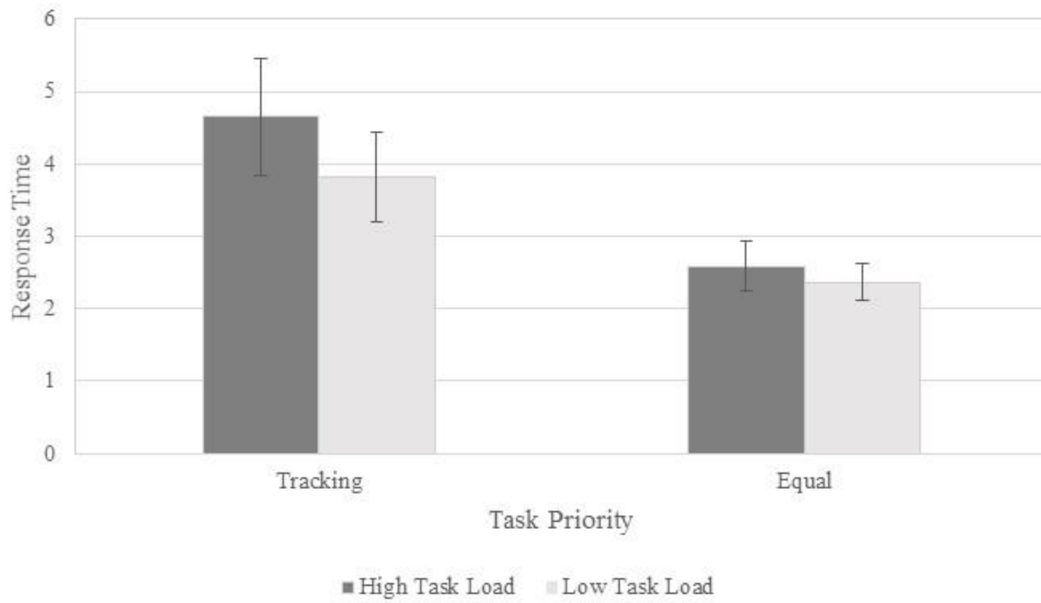

Figure 15. Mean response time to FA events. Error bars represent 95% between-participant confidence intervals.

Hit events. Figure 16 and 17 presents mean response time and mean error rate for Hit events, respectively. The participants responded to Hit events decisively faster in the low task load condition than the high task load condition [M = 2.83 vs. 3.34 seconds; $F(1, 32) = 36.35$, $B_{10} = 8.97 \times 10^3$, $\eta^2_G = 0.11$]. High task priority to the central tracking task however decisively slowed their responses than the equal priority condition [M = 2.42 vs. 3.75 seconds; $F(1, 32) = 31.10$, $B_{10} = 3.04 \times 10^3$, $\eta^2_G = 0.46$]. Data did not provide substantial evidence for the interaction effect [$F(1, 32) = 3.72$, $B_{10} = 1.17$, $\eta^2_G = 0.01$].

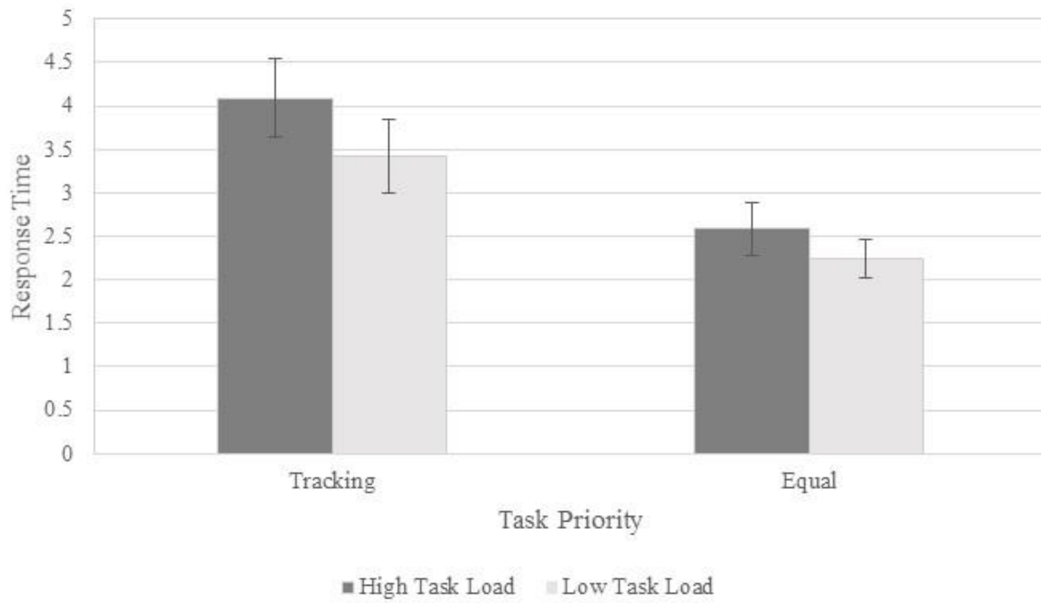

Figure 16. Mean response time to Hit events. Error bars represent 95% between-participant confidence intervals.

For the error rates, the participants committed substantially more errors in the high task load condition than the low task load condition [M = 0.09 vs. 0.06; $F(1, 32) = 6.87$, $B_{10} = 3.46$, $\eta^2_G = 0.03$]. Moreover, error rates in the Hit events were substantially higher in the tracking priority condition than the equal priority condition [M = 0.11 vs. 0.03; $F(1, 32) = 9.70$, $B_{10} = 7.40$, $\eta^2_G = 0.19$]. Data did not produce substantial evidence for the interaction effect [$F(1, 32) = 4.85$, $B_{10} = 1.94$, $\eta^2_G = 0.02$]. Thus, data did not give evidence for the speed-accuracy tradeoffs.

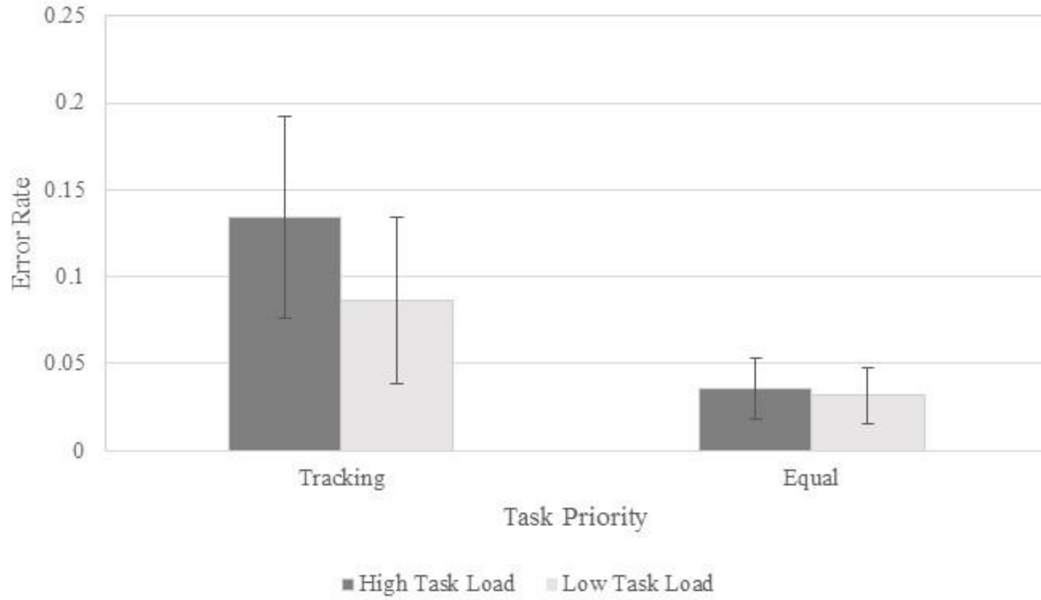

Figure 17. Mean error rate during Hit events. Error bars represent 95% between-participant confidence intervals.

Fuel Management Performance

Figure 18 presents mean fuel for both tanks between task priority conditions. Data provided substantial evidence against the effect of task load for Tank A [$F < 1$, $B_{10} = 1/4.05$] and Tank B [$F < 1$, $B_{10} = 1/3.34$], indicating that the participants maintained similar amounts of fuel regardless of the task load imposed by the central tracking task. Data indicated inconclusive evidence for the main effect of task priority on the amount of fuel in Tank A [M = 2510,94 vs. 2438.80 units; $F(1, 32) = 3.14$, $B_{10} = 1.16$, $\eta^2_G < 0.01$] and Tank B [$F(1, 32) = 1.05$, $B_{10} = 1/1.37$, $\eta^2_G = 0.03$]. Finally, data indicated substantial evidence against the interaction effect on the amount of fuel in Tank A [$F < 1$, $B_{10} = 1/3.23$] while anecdotal evidence against the interaction effect in Tank B [$F < 1$, $B_{10} = 1/2.28$].

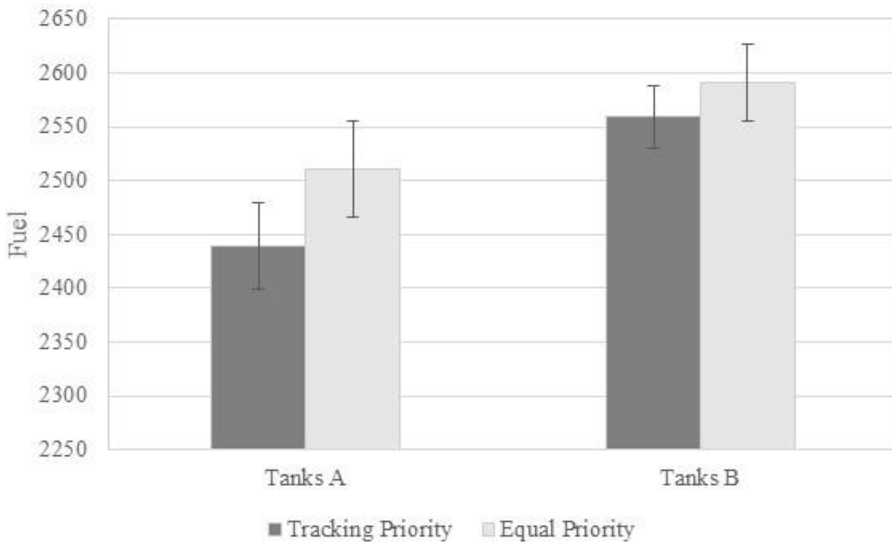

Figure 18. Mean fuel in Tanks A and Tanks B. Error bars represent 95% between-participant confidence intervals.

CHAPTER IV

DISCUSSION

The present study examined whether task priority and task load influence attention allocation and automation trust in a simulated multitasking environment. Using the paradigm used in Karpinsky et al. (2018), participants performed the central tasking task, the system monitoring task with an assistance of an imperfect signaling system, and the resource management task, and their task priority (tracking priority vs. equal priority) and task load (high vs. low on the central tracking task) were manipulated. Though previous works failed to show the effect of task priority on multitasking performance, I aimed to manipulate task priority using the procedure used by Gopher et al. (1982). The purpose of the current study was to replicate the findings of Karpinsky et al. (1982), and examine whether the adverse effects of task load on automation trust was due to task priority placed on the central tracking task.

Comparing with Past Research

Karpinsky et al. (2018) demonstrated that, under higher task load condition, participants 1) scanned the system monitoring task less frequently, 2) rated higher subjective workload, 3) performed the central tracking task more poorly, 4) responded more slowly and made more errors on the system monitoring task, and 5) rated lower levels of performance- and process-based trust. They interpreted the results that participants in the high task load condition misperceived the behaviors of the signaling system due to less frequent information sampling strategies, leading to reduced levels of trust, when compared to the low task load condition, regardless of the reliability level of the signaling system.

The current study generally replicates the findings of Karpinsky et al. (2018). First, participants scanned the system monitoring display less frequently and the central tracking task more frequently in the high task load condition than the low task load condition, supporting Hypothesis 1. Second, participants rated higher subjective workload in the high task load condition than the low task load condition, supporting Hypothesis 2. Third, participants performed the central tracking task more poorly in the high task load condition than the low task load condition, supporting Hypothesis 3. Fourth, participants responded slower to the system monitoring task during FA and Hit events in the high task load condition than the low task load condition. Additionally, participants demonstrated poor system monitoring performance during Hit events in the high task load condition than the low task load condition, supporting Hypothesis 4. The present study found measurable differences in system monitoring performance between the high and low task load conditions perhaps due to the presence of the fuel management task which was absent in Karpinsky et al. (2018). That is, reserve attentional resources were presumably allocated to the fuel management task, increasing difficulty for the system monitoring task. Fifth, participants rated lower levels of performance-based and process-based trust in the high task load condition than the low task load condition, supporting Hypothesis 5. As expected, the participant's subjective workload ratings, MATB performance, and trust ratings were consistent with Karpinsky et al.'s (2018) study, indicating successful replication of Karpinsky et al.'s (2018) findings.

Influence of Task Priority

The current study also aimed to extend Karpinsky et al. (2018) by manipulating task priority in addition to task load. We employed the technique used by Gopher et al.'s (1982) study that successfully manipulated task priority. Later studies, however, failed to manipulate task priority (Gutzwiller et al., 2014; Wickens et al., 2016), thus questioning the importance of task priority in multitasking performance in the literature. The absence of the participant's baseline performance may be one reason why more recent works (Gutzwiller et al., 2014; Wickens et al., 2016) failed to find the effect of task priority on multitasking performance. That is, in Gopher et al.'s (1982) study, participants obtained their own baseline performance by receiving continuous feedback on their tracking performance during the training session. Participants were then instructed to prioritize the tracking task at a priority level of 30%, 50% and 70% of their own baseline performance. Conversely, Gutzwiller et al. (2014) and Wickens et al. (2016) did not provide feedback about the participant's baseline performance for the tracking task prior to the experimental session. Instead, participants were asked to either prioritize the central tracking task or equally prioritize all tasks without providing the participant's baseline performance as an anchor. Additionally, the level of target priority was not explicitly specified in their studies potentially resulting in failed manipulation of task priority.

Using Gopher et al.'s (1982) procedure, participants were instructed to perform the central tracking task at a priority level of either 30% or 70%. Results indicated that participants can equally prioritize all tasks in the high task load condition. Interestingly, participants failed to prioritize the central tracking for more than 70%. Furthermore, in the low task load condition, participants in both priority conditions prioritized the central tracking task more than their target value. Taken together, participants adhered to the instruction when equally prioritizing all tasks in the high task load condition. While participants overprioritized the tracking task in the low load condition and underprioritized in the high load condition, one may wonder whether participants allocated their attention to the system monitoring and the system monitoring tasks appropriately. Indeed, driving research demonstrated increased fixations on a task with high priority, suggesting that task priority can modulate eye movement behavior (Sullivan et al., 2012). Because the current priority manipulation is solely based on the tracking task performance, we do not have a direct way to answer this question.

Previous works (Gopher et al., 1982; Gutzwiller et al., 2014; Wickens et al., 2016) provided contrasting evidence on the effect of task priority. The current study demonstrated the main effect of task priority on attention allocation, system monitoring performance, and system monitoring task. Three points can summarize the results regarding the task priority manipulation. First, participants fixated the system monitoring display more frequently in the equal priority condition than the tracking priority condition, supporting Hypothesis 6. Second, participants responded more slowly to Hit and FA events in the tracking priority condition, compared to the equal priority condition, supporting Hypothesis 7. Additionally, participants committed more errors during Hit events in the equal priority condition than the tracking priority condition. Third, participants performed the central tracking task more poorly in the equal priority condition than the tracking priority condition, supporting Hypothesis 8.

The present study examined the effects of task priority and task load on subjective trust towards the automated system, measured by two separate questionnaires, Jian et al. (2000) and Chancey et al.'s (2017) trust questionnaires. Although reliable evidence was not observed in Jian et al.'s (2000) trust questionnaire, Chancey et al.'s (2017) trust questionnaire revealed notable findings on the effect of task priority. Note that Jian et al.'s (2000) trust questionnaire is empirically driven while Chancey et al.'s (2017) trust questionnaire was theory-driven, and this difference may have caused divergent findings in response to the task priority manipulation. Results showed the interaction effect between task load and task priority on performance-based trust, whereby participants rated lower performance-based trust in the high task load condition than the low load condition when all tasks were equally prioritized, but this effect was eliminated when the tracking task was prioritized. This data pattern is opposite of that predicted by Hypothesis 9. Our tentative interpretation of the result is that participants allocated attention to the tracking task more, and thereby rated their automation trust higher, in the high load condition than the low load condition in the equal priority condition because their attention resources were more mobile. However, when the priority level is set before the experimental session began, then attention resources were less mobile, and the load manipulation became less effective.

The present study offers implications for understanding the relationship between automation trust and attention allocation. Specifically, participants rated lower levels of trust towards the automation when more attention was allocated to the central tracking task. It is possible that participants rated lower levels of trust due to the failure to create a mental model of the automation. Within the HIP model (Wickens et al., 2015), forming an accurate mental model of the automation requires operators to supply sufficient attentional resources to perceive and monitor behaviors of the automated system and maintain the information in working memory. However, operators can fail to form a mental model of the automated system when attentional resources are depleted due to high task load. The current results suggest that both task load and task priority can influence attention allocation strategies, influencing an information uptake for forming an accurate mental model of the automated system.

In Lee and Moray's (1992) framework, trust is based on three informational sources including the performance, process, and purpose. However, later study suggested that novices trust automation based on its performance and process, but not the purpose (Karpinsky et al., 2018). Indeed, task load affected performance-based and process-based trust. The present study demonstrated lower trust ratings of performance- and process-based trust under the high task load condition than the low task load condition, suggesting that operators more occupied with the tracking task developed performance-based trust less. Yet, task priority modulated the effect of task load on performance-based trust, but not the process-based trust nor the purpose-based trust. Based on these findings, equally prioritizing all the tasks likely caused the participants to analyze the automated signaling system's behavior explicitly (e.g., performance-based trust), rather than analyzing the automated signaling system's algorithm (e.g., process-based trust) or intention for developing the automated signaling system (e.g., purpose-based trust).

Practical Implications

The present study demonstrated that high task priority can eliminate the effect of task load on the operator's trust rating and increase attention towards the prioritized task. Practically, the present findings provide insight into designing of training programs involving the use of automation in multitasking environment. Trust has been a critical factor to affect automation use (Parasuraman & Riley, 1997), especially in attention demanding environments where operators exhibit lower ratings of automation trust even though the reliability of the automation remained constant (Karpinsky et al., 2018; Sato et al., 2019). Task priority can be implemented to the training program to control an operator's trust ratings in multitasking environment to encourage appropriate automation use and discourage automation misuse, disuse, and abuse (Parasuraman & Riley, 1997). Training involving task priority has shown dual-task performance benefit for younger and older adults when prioritizing one of the tasks compared to equally prioritizing all tasks (e.g., variable priority training; Kramer et al., 1995). One caveat of the current results is that many of the effect sizes are relatively small and could be of less practical significance. Further research should measure sizes of the effects explored in this study with a larger sample size and more applied environments.

Limitations and Future Study

Several limitations exist in the present study. First, it is unclear whether the current findings can be generalized to experts who have more knowledge than novices. Their expert knowledge may guide trust calibration not solely based on the perception of behaviors of the automation, but process and purpose dimension of automation trust already established from their prior interactions with the automation. Expert operators may also possess an established mental model of the automated system, which could facilitate the trust development process. Future research should examine whether the current findings generalize to expert pilots.

Second, and related to the first point above, it is unclear how Lee and Moray's (1992) trust dimensions develop over time. Previous research on the dynamics of automation trust demonstrated that trust evolved from faith, dependability, and predictability (Muir & Moray, 1996). Although trust dynamics has been examined, none examined the dynamics of Lee and Moray's (1992) trust dimensions chronologically. The present study showed that the experimental manipulations affected mainly performance-based trust. However, the trust questionnaires were administered at the end of each experimental block, reflecting an operator's trust levels only at that time. It is possible that the process- and the purpose-based trust initially increased but decreased towards the end of the experiment. Future research should trace how the three basis of trust develops over time by administering Chancey et al.'s (2017) trust questionnaire at different time points during their interaction with the automated system in the MATB program.

Third, the attractiveness of each task was unknown since the present study did not ask the participants to rate their perceived saliency, interest, difficulty, and task priority attributes of the STOM. For the post-hoc analysis, we used the STOM parameter values reported in Wickens et al. (2016) to analyze the current eye movement data. Future research could collect subjective ratings of the four STOM parameters to compare the model prediction and observed data.

Fourth, the present study did not manipulate risk. Previous studies found risk as an important factor that influences human-automation trust (Chancey et al., 2017; Sato et al., 2019). For example, in Sato et al.'s (2019) study, participants concurrently performed the central tracking task and the system monitoring task where their perceived risk was manipulated. Results indicated that perceived risk magnified the effect of task load, elevating performance- and process-based trust in attention demanding environment. Even though risk is an important factor for trust development, the present study did not consider perceived risk, lowering the ecological validity of the study. Future research should examine how the interaction between perceived risk and task priority influences automation trust.

Fifth, the AOI for each MATB task varied in its size. It is possible that participants made less fixation on the system monitoring display due to the smaller size of the AOI than the others. Future research should consider equating the sizes of the AOIs for the MATB tasks to test this possibility.

Finally, the participants could have used their peripheral vision. In a professional environment where multiple tasks are separated apart, such as gauges and electronic displays within an aircraft cockpit, operators make head movements to direct attention to different tasks that are spatially apart. The present study does not capture this behavior because the participants are required to fixate their head on the chin rest restricting head movement and were able to attend to different tasks by using eye movements. Future research should could a head-mounted eye tracker with multiple displayed spatially apart to examine the effect of information access cost (e.g., Wickens et al., 2015) on attention allocation and automation trust.

CHAPTER V

CONCLUSION

The present study asked participants to perform the MATB paradigm (i.e., central tracking task, system monitoring task, and fuel management task) with varying levels of task load (i.e., low and high) and task priority (i.e., equal priority and tracking priority). Participant were either instructed to perform the central tracking task at a priority level of either 30% (i.e., equal priority) or 70% (i.e., tracking priority). Results demonstrated that participants equally prioritizing all tasks rated lower levels of performance-based trust under high task load condition, consistent with Karpinsky et al. (2018). The present study also showed that the effect of task load in the equal priority condition can be eliminated by setting high priority on the tracking task, suggesting that task priority is one factor that modulates attention allocation and automation trust. The current findings might offer guidelines for devising a training program that can control an operator's trust towards the automation in multitasking environment. By altering the priority of each task, operators can control their trust towards automated systems, which may ultimately allow them to avoiding inappropriate use of automation such as misusing unreliable automation and disusing reliable automation.

REFERENCES

Bailey, N.R., & Scerbo, M.W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science, 8*, 321–348.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775–779.

Barber, B. (1983). *The logic and limits of trust* (Vol. 96). New Brunswick, NJ: Rutgers University Press.

Billings, C.E., (1997). *Aviation automation: The search for a human centered approach.* Mahwah, NJ: Erlbaum.

Bliss, J.P., & Dunn, M.C. (2000). Behavioral implications of alarm mistrust as a function of task workload. *Ergonomics, 43*, 1283–1300.

Breznitz, S. (1984). Cry wolf: The psychology of false alarms. Hillsdale, NJ: Erlbaum.

Chancey, E.T., Bliss, J.P., Yamani, Y., & Handley, H.A.H. (2017). Trust and the compliance reliance paradigm: the effects of risk, error bias, and reliability on trust and dependence. *Human Factors, 57*, 947–958.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Human Factors, 48,* 474-486.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: are they viable? *Human Factors*, *43*, 173-193.

Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays* (NASA Tech, Paper 1711). Moffett Field, CA: National Aeronautics and Space Administration, Ames Research Center.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of experimental psychology: applied*, *1*, 19.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. E. Meyer & S. Kornblum (Eds), *Attention & performance XIV* (*299-322*). Cambridge, MA: MIT Press.

Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 146.

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2014, September). Workload overload modeling: An experiment with MATB II to inform a computational model of task management. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 849-853). Sage CA: Los Angeles, CA: SAGE Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183.

Ishihara, S. (2014). *Ishihara's tests for colour deficiency*. Tokyo: Kanehara Trading Inc.

Jeffreys, H. (1961). *Theory of probability* (*3rd ed.*). New York, NY: Oxford University Press.

Jian, J., Bisantz, A.M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*, 53–71.

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

Karpinsky, N. D., Chancey, E. T., Palmer, D. B., & Yamani, Y. (2018). Automation trust and attention allocation in multitasking workspace. *Applied ergonomics*, *70*, 194-201.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: a comparison of young and old adults. *Journal of experimental psychology: Applied*, *1*, 50.

Lee, J.D, & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics*, *35*, 1243-1270.

Lee, J.D., & See, K.A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*, 50–80.

Li, H., Wickens, C.D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. *Human Factors, 56,* 1050–1061.

Loft, S., Chapman, M., & Smith, R.E. (2016). Reducing prospective memory error and costs in simulated air traffic control: external aids, extending practice, and removing perceived memory requirements. *Journal of Experimental Psychology: Applied, 22*, 272–284.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*, 6-21.

Madsen, M., & Gregor, S. (2000, December). Measuring human-computer trust. In *11th australasian conference on information systems* (Vol. 53, pp. 6-8).

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human factors*, *43*, 563-572.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, *38*, 311-322.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. Boff, L. Kaufman, and Thomas (Eds.), *Handbook of perception and human performance* (pp. 40/1-40/51). New York: Wiley.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*, 1905-1922.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*, 429-460.

National Transportation Safety Board. (2013). *Decent Below Visual Glidepath and Impact with Seawall Asiana Airlines Flight 214 Boeing 777-200ER, HL7742 San Francisco, California July 6, 2013* (Rep. NTSB-AAR1401). Washington, DC: Author.

National Transportation Safety Board (2017). Collisions with Terrain Hageland Aviation Services, Inc. dba Ravn Connect Flight 3153 Cessna 208B, N208SD, Togiak, Alaska October 2, 2016 (Rep. NTSB-AAR1802). Washington, DC: Author.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 91-115). Mahwah, NJ: Erlbaum.

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, *39*, 230-253.

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

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*, 160.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of personality and social psychology*, *49*, 95.

Rice, S. (2009). Examining single- and multiple-process theories of trust in automation. *The Journal of General Psychology*, *136*, 303-322.

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*, 877-903.

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

Safar, J. A., & Turner, C. W. (2005, September). Validation of a two factor structure for system trust. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 3, pp. 497-501). Sage CA: Los Angeles, CA: SAGE Publications.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological review*, *115*, 101.

Santiago-Espada, Y., Myer, R.R., Latorella, K.A., & Comstock, J.R. (2011). The Multi-attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: a User's Guide (NASA/TM-2011–217164). National Aeronautics and Space Administration, Langley Research Center, Hampton, VA.

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2019). Automation trust increases under high workload multitasking scenarios involving risk. *Cognition, Technology & Work*, 1-9.

Senders, J. W. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, 2-5.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Sorkin, R. D. (1988). Why are people turning off our alarms? *The Journal of the Acoustical Society of America*, *84*, 1107-1108.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: a signal detection analysis. *Human Computer Interaction*. *1*, 49-75.

Sullivan, B. T., Johnson, L., Rothkopf, C. A., Ballard, D., & Hayhoe, M. (2012). The role of uncertainty and reward on eye movements in a virtual driving task. *Journal of vision*, *12*, 19-19.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50, 433-441.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*, 291-298.

Wickens, C. D., & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *The International Journal of Aviation Psychology*, *19*, 182-199.

Wickens, C. D., Gutzwiller, R. S., Vieane, A., Clegg, B. A., Sebok, A., & Janes, J. (2016). Time sharing between robotics and process control: Validating a model of attention switching. *Human factors*, *58*, 322-343.

Wickens, C.D., Hollands, J.G., Banbury, S., & Parasuraman, R. (2015). *Engineering Psychology and Human Performance*, Psychology Press.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied*, *1*, 179.

Wickens, C. D., Santamaria, A., & Sebok, A. (2013, September). A computational model of task overload management and task switching. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 763-767). Sage CA: Los Angeles, CA: SAGE Publications.

Yamani, Y., & Horrey, W. J. (2018). A theoretical model of human-automation interaction grounded in resource allocation policy during automated driving. *International Journal of Human Factors and Ergonomics*, *5*, 225-239.

APPENDIX A

INFORMED CONSENT DOCUMENT

PROJECT TITLE: Examining relationships between visual attention and automation trust using eye tracking technique

INTRODUCTION
The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. This research project, *Examining relationships between visual attention and automation trust using eye tracking technique*, will be conducted in Applied Cognitive Performance Laboratory (MGB 325B) at Old Dominion University.

RESEARCHERS
Yusuke Yamani, Ph.D., Assistant Professor, College of Sciences, Department of Psychology, Principal Investigator
Tetsuya Sato, Graduate Student, College of Sciences, Department of Psychology

DESCRIPTION OF RESEARCH STUDY
This research is designed to investigate the ability to perform three concurrent tasks that simulate the control of an aircraft while one of the tasks will be controlled by an automated system with various reliability. We will record both your eye movements and responses during the session. The task will take approximately 2 hours to complete.

EXCLUSIONARY CRITERIA
All participants in this research study must be at least 18 years of age with normal or corrected-to-normal visual acuity and normal color perception.

RISKS AND BENEFITS
RISKS: There may be minimal risk such as eyestrain. The researchers will take all precautions to minimize any of these potential risks.

Eye movements will be monitored by a device that reflects infrared light off the lens and the cornea of the eye. The lens, cornea, and other parts of the eye absorb a small amount of energy from the infrared light, but the energy is less than 1% of the Maximum Permissible Exposure level as certified by the American Standards Institute (ANSI Z 136.1-1973). This is about as much energy you get on a bright sunny day.

BENEFITS: You may not benefit directly from the present study. However, your participation in the study will serve to enhance our understanding of the mechanisms that underlie visual attention.

COSTS AND PAYMENTS
The researchers want your decision about participating in this study to be absolutely voluntary. The main benefit to you for participating in this study is the extra credit or course credit points

that you will earn for your class. Although they are unable to give you payment for participating in this study, if you decide to participate in this study, you will receive 2.5 Psychology Department research credit, which may be applied to course requirements or extra credit in certain Psychology courses. Equivalent credits may be obtained in other ways. You do not have to participate in this study, or any Psychology Department study, to obtain this credit.

CONFIDENTIALITY

The researchers will take reasonable steps to keep private information confidential. The researchers will keep any record of your participation in locked storage in the psychology department. Furthermore, individual participants results will not be distributed in any form. The results of the study aggregated across participants will be published in professional journals and/or book chapters.

WITHDRAWAL PRIVILEGE

It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study at any time. Your decision will not affect your relationship with Old Dominion University, or otherwise cause a loss of benefits to which you might otherwise be entitled.

COMPENSATION FOR ILLNESS AND INJURY

If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact Dr. Yusuke Yamani at 757-683-4457 or Dr. Tancy Vandecar-Burdin the current IRB chair at 757-683 3802 (or at tvandeca@odu.edu) at Old Dominion University, or the Old Dominion University Office of Research at 757-683-3460 who will be glad to review the matter with you.

VOLUNTARY CONSENT

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

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

And importantly, by verbally agreeing, you are telling the researcher that you DO agree to participate in this study.

APPENDIX B

DEMONGRAPHICS FORM

Demographic Information Sheet
Applied Cognitive Performance Laboratory

Office use:	ACP ID: _________
Near Vision: _______ Far: ______ Color: _____	

Name:
____

Date of Birth: ______________ Age: _____

Health: 1 2 3 4 5
Poor → Excellent (circle one)

Gender: ☐Male ☐Female

Race: ______________________________

Native Language: ____________________ Second Language: ______________________

Please Circle True/False for the following.

Do you wear Glasses/Contacts on a regular basis? True False
Have you been diagnosed with any neuropsychological dysfunction? True False
If so, are you currently taking any medication for this? True False

How many years of education have you completed (Please record a number)? _____
Please note: grade school through high school is usually 12 years in the US, if needed,
Add on how many years of college you have completed.

Contact information
*Home Phone number :___(____)_______________ *Cellular :___(____)_______________

*Email: ____

*Address: ____

*If you are interested in receiving information about future experiments, please include address and please answer below. Your address, phone number and email will not be used for any other purpose.
Can we contact you to participate in additional paid experiment? Yes No

Where did you hear about us? ____________________________________

Signature of Participant: ____

Name (please print):____

APPENDIX C

NASA-TASK LOAD INDEX (TLX) WORKLOAD QUESTIONNAIRE

(Hart & Staveland, 1998)

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses workload on five 7-point scales. Increments of high, medium, and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Very Low — Very High

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low — Very High

APPENDIX D

TRUST QUESTIONNAIRE (Chancey et al., 2017)

Part. #: ________________ Group: ___________________ Session: ___________________

Below is a list of statements for evaluating trust between people and automated systems. Please circle the number that best describes your feeling or your impression of the automated aid you used during the task.

1. Even when the automated aid gives me unusual advice, I am certain that the aid's advice will help me to perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

2. For me to perform well, I can rely on the automated aid to function properly

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

3. It is easy to follow what the automated aid does to help me perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

4. The automated aid's advice reliably helps me perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

5. The automated aid's advice consistently helps me perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

6. I understand how the automated aid will help me perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

7. Even if I have no reason to expect that the automated aid will function properly, I still feel certain that it will help me to perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

8. Although I may not know exactly how the automated aid works, I know how to use it to perform well.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

9. To help me perform well, I believe advice from the automated aid even when I don't know for certain that it is correct.

   Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

10. To help me perform well, I recognize what I should do to get the advice I need from the automated aid the next time I use it.

    Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

11. I will be able to perform well the next time I use the automated aid because I understand how it behaves.

    Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

12. The automated aid always provides the advice I require to help me perform well.

    Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

13. The automated aid adequately analyzes the system consistently, to help me perform well.

    Not Descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 :Very Descriptive

APPENDIX E

TRUST QUESTIONNAIRE (Jian et al., 2000)

Checklist for Trust between People and Automation

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark an "x" on each line at the point which best describes your feeling or your impression.

(Note: not at all = 1: extremely = 7)

1. The system is deceptive

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

2. The system behaves in an underhanded manner

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

3. I am suspicious of the system's intent, action, or outputs

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

4. I am wary of the system

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

5. The system's actions will have a harmful or injurious outcome

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

6. I am confident in the system

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

7. The system provides security

|_______|_______|_______|_______|_______|_______|_______|
1 2 3 4 5 6 7

8. The system has integrity

|_________|_________|_________|_________|_________|_________|_________|
1 2 3 4 5 6 7

9. The system is dependable

|_________|_________|_________|_________|_________|_________|_________|
1 2 3 4 5 6 7

10. The system is reliable

|_________|_________|_________|_________|_________|_________|_________|
1 2 3 4 5 6 7

11. I can trust the system

|_________|_________|_________|_________|_________|_________|_________|
1 2 3 4 5 6 7

12. I am familiar with the system

|_________|_________|_________|_________|_________|_________|_________|
1 2 3 4 5 6 7

VITA

Tetsuya Sato
Old Dominion University
Mills Godwin Building 325B
Norfolk, VA 23529

Education

Master of Science in Psychology, Human Factors concentration | August 2020
Old Dominion University, Norfolk, VA
Advisor: Yusuke Yamani, Ph.D.

Bachelor of Sciences in Psychology, Human Services minor | May 2018
Old Dominion University, Norfolk, VA

Publication

Yamani, Y., Chancey, E. T., Xu, V., & **Sato, T.** (in preparation). Validation of the Performance, Process, Purpose (P^3) human-automation trust questionnaire.

Sato, T., Yamani, Y., & Millner, N. (in preparation). Frequent auditory interruption does not impact the effect of visual task load on trust towards imperfect automation: A flight simulator study.

Long, S., **Sato, T.,** Millner, N., Mirabelli, J., Loranger, R., & Yamani, Y. (under review). Empirically and theoretically driven scales on automation trust: A multi-level confirmatory factor analysis. *To be presented at the Human Factors and Ergonomics Society 2020 Annual Meeting.*

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2019). Automation Trust Increases Under High Workload Multitasking Scenarios Involving Risk. *Cognition, Technology & Work,* 1-9.

Tremmel, C., Herff, C., **Sato, T.**, Rechowicz, K., Yamani, Y., & Krusienski, D. (2019). Estimating Cognitive Workload in an Interactive Virtual Reality Environment using EEG. *Frontiers in Human Neuroscience*, *13*, 401.

Research Experience

Research Assistant | University of Tsukuba (July 2019 – August 2019)
Supervisor: Makoto Itoh, Ph.D.

Graduate Research Assistant | Old Dominion University (June 2018 - Present)
Supervisor: Yusuke Yamani, Ph.D.

Undergraduate Research Assistant | Old Dominion University (2017-2018)
Supervisor: Yusuke Yamani, Ph.D.