Summer 2019

The Effects of Automation Transparency and Reliability on Task Shedding and Operator Trust

William Everett Lehman
Old Dominion University, welehman@gmail.com

THE EFFECTS OF AUTOMATION TRANSPARENCY AND

RELIABILITY ON TASK SHEDDING AND OPERATOR TRUST

by

William Everett Lehman
B.S. May 2010, Michigan Technological 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 2019

Approved by:

James P. Bliss (Director)

Jing Chen (Member)

Matt R. Judah (Member)

ABSTRACT

THE EFFECTS OF AUTOMATION TRANSPARENCY AND RELIABILITY ON TASK SHEDDING AND OPERATOR TRUST

William Everett Lehman
Old Dominion University, 2019
Director: Dr. James P. Bliss

Because automation use is common in many domains, understanding how to design it to optimize human-automation system performance is vital. Well-calibrated trust ensures good performance when using imperfect automation. Two factors that may jointly affect trust calibration are automation transparency and perceived reliability. Transparency information that explains automated processes and analyses to the operator may help the operator choose appropriate times to shed task control to automation. Because operator trust is positively correlated with automation use, behaviors such as task shedding to automation can indicate the presence of trust. This study used a 2 (reliability; between) × 3 (transparency; within) split-plot design to study the effects that reliability and amount of transparency information have on operators' subjective trust and task shedding behaviors. Results showed a significant effect of reliability on trust, in which high reliability resulted in more trust. There was no effect of transparency on trust. There was no effect of either reliability or transparency on task shedding frequency or time to task shed. This may be due to high workload of the primary task, restricting participants' ability to utilize transparency information beyond the automation recommendation. Another influence on these findings was participant hesitance to task shed which could have influenced behavior regardless of automation reliability. These findings contribute to the understanding of automation trust and operator task shedding behavior. Consistent with

literature, reliability increased trust. However, there was no effect of transparency, demonstrating the complexity of the relationship between transparency and trust. Participants demonstrated a bias to retain personal control, even with highly reliable automation and at the cost of time-out errors. Future research should examine the relationship between workload and transparency and the influence of task importance on task shedding.

This thesis is dedicated to my family for telling me how smart and handsome I am. And also for their constant support of my academic endeavors. This is also for my friends who provided unconditional support, even when it should have been conditional, Zach, Kerri, and Morgan.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Bliss, for his guidance in my own research as well as the other research and work opportunities he has contributed to along the way. A special thanks to my fantastic committee, Dr. Chen and Dr. Judah, who provided feedback to make this thesis academically sound. Alex, Catherine, and Shelby, you all have been incredible labmates. From commiseration over programming to all of the times you have put up with my “hey, do you have a second for a question.” Thank you, Seterra, for allowing me to talk at you about automation until I figured out what I was trying to say. And of course, everyone else at ODU that has been involved in this process.

I will be forever grateful for my friends and faculty from Michigan Tech who helped prepare me for this. Susie whose advising contributed to my grad school decisions, Kerri who showed me how to do it all and still succeed, and Jinan for her passion for research.

And finally, thank you to the Fairgrounds staff who provided the caffeine for hours and hours of reading, writing, procrastinating, and rewriting.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Automation use can aid operators in complex work environments. By taking over processes and decisions from the operator, automation provides the benefits of improved human-automation performance and reduced operator workload (Parasuraman & Riley, 1997). Trust is an important factor for establishing appropriate automation use (Lee & See, 2004). Specifically, calibrated trust that corresponds to automation capabilities helps operators make more efficient automation use choices with fewer errors (Lee & See, 2004; Parasuraman & Riley, 1997). Transparency is an automation characteristic that provides the operator with explanation of automated processes (Lyons, 2013). This explanation helps operators calibrate trust by facilitating system appraisal (Barnes, Chen, & Hill, 2017; Chen et al., 2014). Automation reliability affects trust: increasing reliability leads to increased trust and more use of automation (de Visser & Parasuraman, 2011; Hancock et al., 2011; Ma & Kaber, 2007).

The goal of the proposed study was to examine the combined effects of automation transparency and reliability on operators' self-reported trust as well as task shedding behaviors in the Information, Surveillance, and Reconnaissance (ISR) domain. Robot performance-based factors, which includes transparency and reliability, have a larger effect on trust development than human-related, robot attribute-based, or environmental factors (Hancock et al., 2011). Because of this, it is important to examine how transparency and reliability can have a combined influence on trust and automation use. Trust that is accurately calibrated to automation reliability is important for good human-automation system performance. Therefore, understanding how transparency can interact with different levels of reliability is important for automation design.

Military ISR requires information assembly, analysis, and interpretation, tasks that could benefit from automation (Adams, Bruyn, Houde, & Angelopoulos, 2003; Tyler, 1999). Implemented judiciously, automated decision aids in ISR could improve operator performance and reduce cognitive workload. Effective trust calibration happens when an ISR operator has an accurate understanding of automation strengths and weaknesses. Operators may then choose to use automation, and in some cases, may entirely shed a task to automation.

Automation

Automation is the use of technology to accomplish tasks that had previously been accomplished by a human (Madhavan & Wiegmann, 2007; Parasuraman & Riley, 1997). Lee and See (2004) characterized automation as technology that selects data, transforms information, makes decisions, or controls processes. Use of automation provides a range of benefits, depending on the situation. Decision aids can quickly analyze and compute information, and teleoperated automation can remove human workers from dangerous environments. Automation can also accomplish tedious tasks without tiring or losing attentional focus or can complete tasks that humans are not physically able to do such as lifting heavy equipment (Adams et al., 2003).

One example of automation use in ISR is synthetic vision which is the use of augmented reality (AR) with see-through head mounted displays (HMD). Synthetic vision provides operators with direct view of physical terrain along with overlaid AR text, icons, or models of occluded terrain (Foyle, Ahumada, Larimer, & Sweet, 1992; Livingston et al., 2002; Livingston et al., 2003). Synthetic vision can be used in a range of applications, one of which is the urban battlefield. In systems using synthetic vision, automation selects data from the terrain or sensors and transforms those data into a visual representation, potentially even suggesting or choosing routes or making tactical decisions depending on automation level.

Levels of automation (LOAs) are characterized by the amount of human and automation contribution to system decision selection and action implementation (Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978). At the lowest LOA, the operator makes decisions and implements actions without automation contribution. At the highest LOA, automation makes decisions and acts autonomously. Trust in automation is an important factor contributing to operator use of automation (Lee & See, 2004; Parasuraman & Riley, 1997). Because higher LOAs feature more automated task control, trust is more important for determining automation use. Calibrating trust to the given LOA increases safety, efficiency, and human-automation system productivity by reducing automation use errors (Lee & See, 2004; Muir, 1994).

Trust

Operators are more likely to use automation when there is trust that the automation will benefit operator goal attainment (Endsley, 2017; Lee & See, 2004; Parasuraman & Riley, 1997). Mayer, Davis, and Schoorman (1995) defined trust as the trustor's willingness to be vulnerable to a trustee's actions, expecting that the trustee will act in a way important to the trustor. This willingness holds even when the trustor has no control over the trustee's behavior. Mayer et al.'s (1995) social trust model comprises the trustor's natural propensity to trust and the perceived trustworthiness of the trustee (see Figure 1). The factors that impact trustworthiness are perceived ability, benevolence, and integrity (Mayer & Davis, 1999). The trustee's ability includes skills, competencies, and characteristics that facilitate influence in a certain domain. Benevolence is the trustee's desire to do good for the trustor regardless of personal gain. Finally, integrity is the expectation that the trustee will act according to guidelines acceptable to the trustor. These three factors vary independently and combine with the trustor's propensity to trust resulting in social trust of a trustee (Mayer et al., 1995).

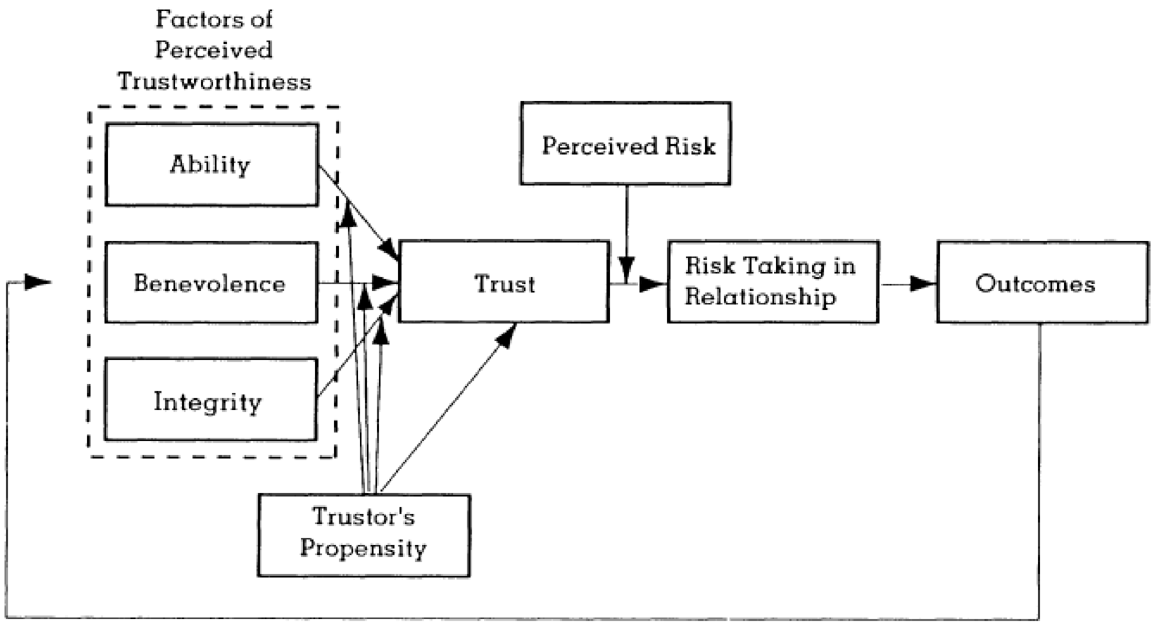


Figure 1. Model of social trust proposed by Mayer et al. (1995).

Building on social trust, researchers have argued for the similarity of social trust to trust in automation (Adams et al., 2003; Lee & See, 2004; Muir, 1994). Lee and See (2004) defined trust in automation as an attitude toward automation that, in situations of uncertainty and vulnerability, the automation will help achieve the operator's goals. This definition includes three bases (performance, purpose, and process) that are comparable to Mayer et al.'s (1995) ability, benevolence, and integrity, respectively. Performance includes the reliability, predictability, and capability of the automated system. Purpose describes automation use that follows the designer's intentions. Process is the appropriateness of using automation for a given task. Trust is then developed along the dimensions of calibration, resolution, and sensitivity. Operators calibrate trust in automation by coordinating trust to the demonstrated capabilities of the automation. Resolution refers to the operator's ability to differentiate among LOAs. Sensitivity is the influence of a specific automation characteristic on trust (Lee & See, 2004).

Two characteristics of automation that affect operator trust are transparency and reliability. Automation reliability refers to the proportional accuracy of the automated decision aid's recommendations. More reliable automation leads to increased operator trust and use of automation (Chavaillaz, Wastell, & Sauer, 2016; de Visser & Parasuraman, 2011; Hancock et al., 2011; Ma & Kaber, 2007). Transparency is the characteristic of a system that communicates information about automated processes and the current state of the automation (Chen et al., 2014; Lyons, 2013; Ososky, Sanders, Jentsch, Hancock, & Chen, 2014).

Transparency

Transparency has been described as *seeing into* or *seeing through* a system (Ososky et al., 2014; Sheridan & Verplank, 1978). Operators who "see through" a system feel as though they are directly manipulating their target without the intervening automation. This type of system transparency is meant to provide direct perception of the target's state rather than the automation's state. Operators who "see into" the system receive information regarding how and why system processes are proceeding in the current state (Ososky et al., 2014; Sheridan & Verplank, 1978). This information is used to provide operators with information about automated processes, analyses, recommendations, and actions made by the observable automated decision aid.

Two models of transparency involve situation awareness and human-robot information exchange. The situation awareness-based agent transparency model developed by Chen and colleagues (2014) describes levels of transparency based on Endsley's (1995) three levels of situation awareness (SA). Level 1 transparency communicates the automated system's plans, goals, and current state. Level 2 communicates automation plans and actions. Level 3 communicates automation projection of future actions (Chen et al., 2014).

Lyons' (2013) transparency taxonomy includes robot-to-human and robot-of-human factors split into individual models. In developing this taxonomy, Lyons uses "robot" and "automation" interchangeably. Robot-of-human transparency communicates the automation's understanding of operator state to the operator. Robot-of-human transparency comprises the Teamwork Model and the Human State Model. The Teamwork Model conveys the robot's knowledge of human-robot responsibility sharing and current autonomy level. In the Human State Model, the robot communicates awareness of the human's cognitive, physical, and emotional states.

Robot-to-human transparency describes the automation communicating understanding of its own abilities, intentions, and situational constraints to the operator (Lyons, 2013). Robot-to-human transparency includes Intentional, Task, Analytical, and Environment models (Lyons, 2013). Transparency in the Intentional Model communicates information about the purpose of the robot, how it will behave to fulfill this purpose, and the framework of how the robot is programmed to interact with humans. The Task Model can communicate the information regarding the robot's current goal and progress toward that goal, as well as communication that the robot is aware of its own capabilities and any errors made. The Environment Model should communicate the robot's understanding of how terrain and weather conditions affect function, the potential for hostile interaction, and ability to switch between high and low demand functionality. The Analytical Model communicates the analytical processes that underlie the robot's decision-making process. This can include how information is combined from multiple sources and the underlying logic explaining how a robot arrived at a decision (Lyons, 2013).

Trust Calibration

Transparency and reliability jointly affect trust calibration and automation use. Calibrated trust allows the operator to rely on a trustworthy system and direct attention to less reliable systems (Muir, 1987). Calibration can be facilitated by providing operators with transparency information communicating system uncertainty, automation state, limitations, or capabilities (Helldin, Falkman, Riveiro, Dahlborn, & Lebram, 2013; McGuirl & Sarter, 2006; Merlo, Wickens, & Yeh, 1999). Conveying how specific environmental factors affect automation capabilities also improves trust calibration (Lee & See, 2004). However, if processing raw transparency information is difficult for the operator, the transparency may hinder calibration (Wickens, Gempler, & Morphew, 2000). Given complex raw data, consolidating individual pieces into a more comprehensive explanation may be beneficial to operator cognitive workload (Lyons, 2013).

Reliability also influences trust calibration. Automation reliability is positively related to operator trust and use of automation (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Hancock et al., 2011; Ma & Kaber, 2007). Wiegmann, Rich, and Zhang (2001) found that operators underestimate automation reliability, but that reliability estimates become more accurate over time. As operators lose trust following automation errors, recalibration must occur to regain trust. By interacting with properly functioning automation, trust can be recovered over time (Lee & Moray, 1992; Merlo, 1999). Providing operators with explicit reliability information helps trust calibration occur more quickly and accurately (Lee & Moray, 1994; Merlo, 1999; Wang, Jamieson, & Hollands, 2009).

Automation Use

From the above discussion, trust is an important influence on automation use decisions. Meyer (2001) described two automation use behaviors. Compliance is operator agreement with actions endorsed by automation. Reliance occurs when the operator does no action, accepting that the lack of an automated signal accurately indicates lack of a problem (Meyer, 2001). Trust is positively correlated with compliance and reliance (Lee & Moray, 1992; Muir, 1994). As trust in automation declines, so does automation use, implying that trust precedes automation use (Moray & Inagaki, 1999). Ultimately, an operator’s choice to use automation can indicate the presence of trust in that automation (Boubin, Rusnock, & Bindewald, 2017; Lee & See, 2004; Parasuraman & Riley, 1997).

Though trust often precedes automation use, the two are not perfectly correlated (Ma & Kaber, 2007; Wiegmann et al., 2001). Some operators make automation use decisions prior to actual use opportunity (Bliss, Harden, & Dischinger, 2013). These decisions indicate that the operator has formulated an automation use strategy without considering trust calibrated over time. Wang, Pynadath, and Hill (2016) found that, although 100% reliable automation should be entirely trustworthy, operators may still demonstrate less than 100% compliance.

On the contrary, there are situations where operators demonstrate automation use without trust. Mandated use of automation or task overload may result in use of untrusted or only marginally reliable automation (Bliss & Gilson, 1998; Chancey, Proaps, & Bliss, 2013; Rice, 2009). Another situation in which trust does not precede automation use occurs when the operator’s self-confidence is low. In this case, the operator may use automation that is untrustworthy, but is still believed to be better than the operator’s abilities (Wiegmann et al., 2001).

One challenge in designing automation is facilitating appropriate use behaviors to help the operator avoid misuse, disuse, and abuse errors (Lee & See, 2004). Misuse is operator use of unreliable automation. Such behavior can result in errors and a loss of operator SA, creating potential for danger. Disuse occurs when the operator does not utilize reliable automation that would benefit human-automation system performance. Abuse refers to operator use of automation in ways not intended by its designers (Lee & See, 2004). Establishing proper use of automation is important for effective human-automation system performance.

Transparency and automation use. Operators are more likely to utilize automation that is transparent (Helldin et al., 2013; Lyons, 2013; Ososky et al., 2014). Transparency helps operators calibrate their mental model of automation. An accurate mental model makes the decision to use automation easier by reducing the cognitive overhead associated with the use decision (Parasuraman & Riley, 1997). Having an accurate mental model also helps the human-automation system more readily benefit from automation use while reducing misuse, disuse, and abuse behaviors.

Transparency not only increases automation use, but also improves performance when the operator chooses to use automation (Barnes et al., 2017; Ososky, et al., 2014). Transparency can facilitate trust calibration to automation capabilities, helping operators reduce misuse and disuse errors (Barnes et al., 2017; Vincente & Rasmussen, 1990). Providing an automation decision recommendation with rationale transparency can reduce operator performance errors compared to providing no rationale or providing a rationale and timestamp (Wright, Chen, Barnes, & Hancock, 2017). Transparent information exchange between the operator and automation communicating understanding of each other’s overall abilities, performance, and current state may improve human-automation system performance (Chen et al., 2014; Rouse, 1994; Scerbo, 1994).

Reliability and automation use. Increased reliability also leads to greater operator trust and use of automation (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Hancock et al., 2011; Helldin et al., 2013; Ma & Kaber, 2007; Ososky et al., 2014; Parasuraman & Riley, 1997). In general, higher reliability improves human-automation system performance (Chavaillaz et al., 2016; Rovira, McGarry, & Parasuraman, 2007; Wickens, Dixon, & Ambinder, 2006). However, operators are less likely to monitor automation performance in high reliability conditions and are then more likely to miss when the automation does make errors. In low reliability conditions, operators monitor automation performance more and are less likely to miss an automation error (Endsley, 2017). Wickens and colleagues (2006) found that, when automation does make errors, false alarms lead to worse performance than misses.

When the operator has confidence in his or her ability to successfully complete a task without automation, they are less likely to use automation (Chavaillaz et al., 2016; Madhavan & Wiegmann, 2007; Moray, Inagaki, & Itoh, 2000; Parasuraman & Riley, 1997). This effect of self-confidence is more pronounced when system reliability is low. In such cases, the operator may have more self-confidence to complete the task without automation because of the unreliability of the automation (de Visser & Parasuraman, 2011). When automation completes a task the same way an operator would complete it, the operator may be more likely to use that automation (Boubin et al., 2017). This could be due in part to better operator understanding of automation processes. On the contrary, system complexity may make it more difficult for the operator to develop an accurate mental model of automation function, reducing automation use (Endsley, 2017).

Task Shedding

Automation design impacts use of automation, specifically whether automation is dynamic or static. Static automation is consistent in capabilities or LOA throughout a task. Dynamic automation flexibly adjusts to operator use or LOA (Parasuraman & Hancock, 2001). Dynamic automation is further classified as adaptive or adaptable, depending on whether the operator or the automation is responsible for task allocation (Parasuraman & Wickens, 2008). Adaptive automation is characterized by task allocation initiated by automation (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992). Conversely, adaptable automation is characterized by task allocation initiated by the operator (Parasuraman & Wickens, 2008). The operator may choose LOA or may choose to retain control or allocate a task to automation. Relinquishing task control is also known as adaptive task allocation to the machine (ATA-M) or task shedding (Chavaillaz et al., 2016; Parasuraman & Hancock, 2001). The benefits of operator control of task shedding include decreasing operator workload and improving human-automation system performance, either by increasing reliance on highly reliable automation or by decreasing reliance on unreliable automation (Parasuraman & Hancock, 2011; Parasuraman, Mouloua, & Hilburn, 1999). Some researchers have suggested that adaptive automation can hurt performance by increasing system unpredictability associated with automated task responsibility changes (Billings & Woods, 1994). However, allowing operators to control task shedding may mitigate the negative effects of this unpredictability (Miller & Parasuraman, 2007; Parasuraman, Galster, Squire, Furukawa, & Miller, 2005).

Operator trust and automation reliability, transparency, and design can impact general automation use. There is evidence that these and certain other factors specifically affect task shedding frequency. Bliss and colleagues (2013) found that operators are more likely to task shed control to automation that has demonstrated high reliability. Operators are also more likely to task shed in situations of high workload or low certainty, which may serve to reduce the cognitive workload associated with uncertainty (Parasuraman & Hancock, 2001).

Two factors that may reduce operator tendency to task shed are high cognitive overhead associated with task shedding decisions and low trust in automation (Hancock et al., 2011; Parasuraman & Riley, 1997). With an accurate mental model of automation facilitated by transparency, the operator's decision to task shed or retain control should be quicker and lead to more efficient human-automation task sharing. Other aspects of cognitive overhead include the time required to engage automation and the opportunity costs of doing so (Parasuraman & Riley, 1997). Hancock and colleagues (2011) found that when operators have less trust in automation, they assume control sooner. This again demonstrates automation use as an indicator of trust.

Two operator biases in task shedding have been observed: retention of personal task control and immediate task shedding (Bliss et al., 2013; Lee & Moray, 1994; Parasuraman & Riley, 1997). Retention of task control increases when operators believe they can succeed without the aid of automation. However, operators also favor the status quo. This means that operators in control tend to retain control, but once control is given to automation, the operator tends to leave control to the automation (Parasuraman & Riley, 1997). Bliss and colleagues (2013) observed that, of the participants who task shed, many chose to immediately shed control to automation, indicating a task shedding decision prior to trial participation.

Current Study

Although both reliability and transparency influence automation use few researchers have examined how these factors interact in operator use of automation. Kaltenbach and Dolgov (2017) examined the effects of automation reliability and transparency on operator trust when interacting with the automated Coffee-O-Matic interface. The Coffee-O-Matic interface uses a simulated coffee production task in which operators maintain fluctuating temperature and pressure states within a target range. Reliability was operationalized as whether or not the automation executed operator input. Transparency provided information about the automated process that was occurring. Amount of transparency was manipulated by displaying either the current process or the current process along with historical information of previous processes. They found a significant effect of transparency when measuring trust with the Trust in Automated Systems Scale, but no effect of reliability. They found no effect of transparency or reliability when measuring trust with the Human Computer Trust Scale. This research should be expanded to examine how the amount of transparent information regarding current processes and the reliability of those automated processes influence operator trust and automation use choices.

To date, research on automation has not addressed how amount of transparency and reliability may jointly influence the human-automation system. Because operators may interact with automated systems that differ in reliability, it is important to understand how transparency can influence task shedding decisions under different reliability conditions. The goal of the proposed study was to examine the joint effects of automated decision aid transparency and reliability on subjective trust and task shedding behavior during a simulated ISR task. The proposed study examined transparency through Lyons' (2013) Analytical Model which communicated analysis and rationale underlying automated processes. Amount of transparency was manipulated by displaying different amounts of information, all about current automated processes. Reliability was examined through the accuracy of automated decision recommendations and supporting analysis. Because the reliability of automated decision aids often is not perfect, research was needed to understand how transparency should best communicate system information to facilitate optimal automation use behaviors.

Hypotheses

H1. For subjective trust, automated system transparency and reliability were predicted to interact. Trust was expected to be similar across transparency conditions that were highly reliable. However, for low reliability systems, greater transparency was expected to result in less trust. Subjective trust at high reliability was expected to be similar for high and low transparency (see Figure 2). This hypothesis is derived from Kaltenbach and Dolgov's (2017) study reflecting results using positive valence questions from the Trust in Automated Systems Scale. Their results showed that increasing transparency had a negative effect on trust in the low reliability condition, but did not have an effect in the high reliability condition.

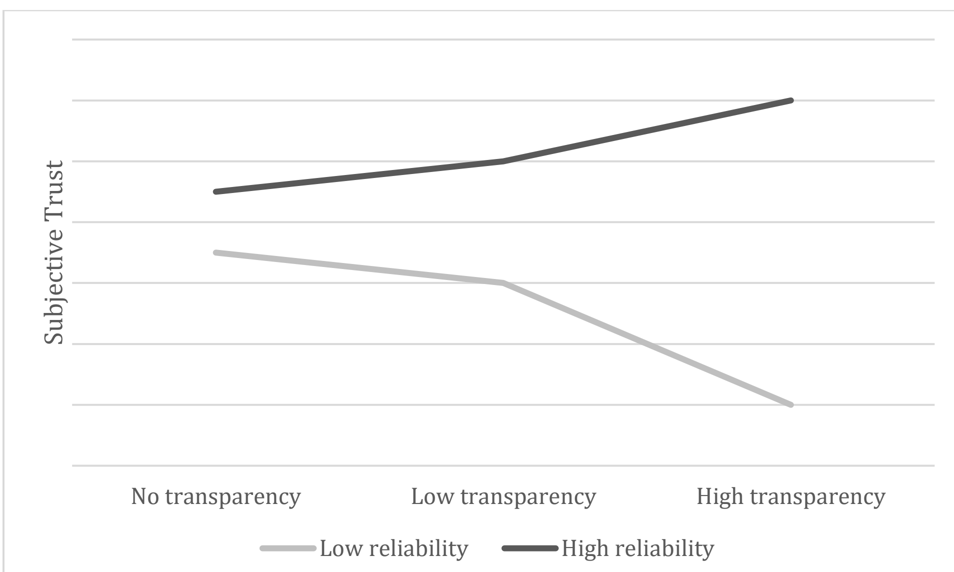


Figure 2. Predicted interaction of transparency and reliability on operator trust.

H2. For percentage of time the task is shed, automated system transparency and reliability were predicted to interact. Participants were expected to task shed most when system activities were transparent and highly reliable. They were expected to task shed least when system activities were less reliable and highly transparent (see Figure 3). This hypothesis is derived from findings from Barnes and colleagues (2017) that misuse and disuse errors with automation decreased with high transparency. These findings indicate that high transparency would increase frequency of task shedding to reliable automation, thereby reducing disuse. It also indicates that high transparency would decrease frequency of task shedding to unreliable automation, thereby reducing misuse.

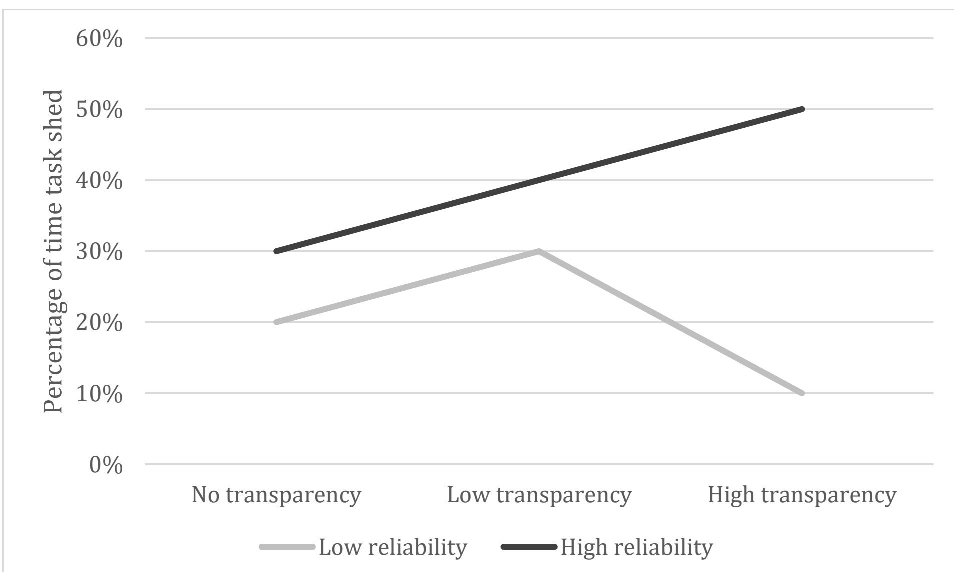


Figure 3. Predicted interaction of transparency and reliability on percentage of time task shed.

H3. We expected to find a main effect of transparency information in which more transparency would evoke more frequent operator task shedding. This would be reflected in a significant *F* test of effect of transparency information across levels of reliability. This hypothesis was derived from research indicating that transparency increases operator understanding of a system (Chen et al., 2017) and willingness to use that system (Helldin et al., 2013; Ososky et al., 2014).

H4. The high reliability condition, compared to the low reliability condition, was expected to evoke more frequent task shedding. This would be supported by a significant *F* test, comparing reliability conditions across levels of transparency. Increasing reliability has been demonstrated to increase use of automation (Helldin et al., 2013; Ososky et al., 2014; Parasuraman & Riley, 1997).

H5. Increased reliability was expected to cause more rapid task shedding. This would be supported with a significant *F* test which compares levels of reliability collapsing across levels of transparency. Having more reliable automation can make the decision to use automation easier, leading to faster automation use decisions (Bliss et al., 2013; Ososky, et al., 2014).

CHAPTER 2

METHOD

Design

This experiment employed a 2 (reliability; 60% or 90%) × 3 (transparency; none, low, or high) split-plot design. Participants were engaged in a primary tracking task using the Multi-Attribute Task Battery II (MATB-II) along with a tank spotting task (Santiago-Espada, Myer, Latorella, & Comstock, 2011). The tank spotting task was designed to simulate an ISR task in which participants make target absence or presence judgements with the help of an automated decision aid. Chancey (2016) used a similar task to study trust and operator interaction with automated systems. The automated decision aid for the secondary task was used to manipulate levels of reliability and transparency while participants performed the tank spotting task with the option to task shed the secondary task to the automation. Used together, these tasks were meant to simulate a multi-task environment in which the operator focuses on simulated flight tasks while performing ISR-type target identification.

Independent Variables. Level of reliability and amount of information communicating system transparency were manipulated by modifying the actions of an automated decision aid for the tank spotting task. Transparency was operationalized as the amount of information provided to participants explaining how automation processes were proceeding and why recommendations were made. This method of operationalization is commonly used in transparency research (Chen, Barnes, Wright, Stowers, & Lakhmani, 2017; Helldin et al., 2013; Kaltenbach & Dolgov, 2017). Amount of information was manipulated as a within-subjects variable at three levels: no information, low information, and high information. Transparency was manipulated according to recommendations within Lyons' (2013) Analytical Model: transparency information communicates analysis and process information that precedes automation decision making. For each level of transparency, the participant was presented an automation judgement of "Tank Present" or "Tank Absent" followed by information about how the recommendation was made (i.e. what analysis) and why the recommendation was made (i.e. detected values exceeded detection threshold). Information was available for two attributes: vehicle information and traffic information. In the no transparency condition, only the decision recommendation was presented. In the low transparency condition, the attributes displayed a single analysis and why the recommendation was made. High transparency information displayed two analyses and two reasons why the recommendation was made (see Appendix A).

Manipulating reliability between subjects served two purposes. First, it reduced the chance of carry-over effects from changing reliability, thereby facilitating operator calibration (Wiegmann, Rich, & Zhang, 2001). Also, the between-subjects manipulation maintained a shorter testing session. Past research has shown performance effects in conditions of varying automation reliability. There is evidence that use of automation that is less than 70% reliable harms human-automation system performance compared to performance of the operator alone (Chavaillaz et al., 2016; Rovira et al., 2007; Wickens et al., 2006). Other research has suggested performance decrements with automation reliability as high as 80% or 90% (Hillesheim & Rusnock, 2016; Moray et al., 2000; Scerbo, 1996; Wickens et al., 2006). The 60% low reliability condition was below these suggested thresholds at which the use of the automation may hurt human-automation performance compared to sole operator performance. The 90% high reliability was expected to elevate human-automation performance beyond unaided operator performance. These reliability levels mirrored those used in other research on the effects of transparency or reliability on operator trust and automation use behaviors (Chancey, Bliss, Proaps, & Madhavan, 2015; Kaltenbach & Dolgov, 2017). The historical reliability of the automation was communicated to participants through a vignette at the beginning of the study (see Appendix B). Communicating reliability helps stabilize performance to minimize variability associated with trust calibration (Helldin et al., 2013; Wang et al., 2009).

Dependent Variables. Measures of the dependent variables were taken from a subjective trust questionnaire as well as task shedding behaviors. Subjective trust was measured with an adapted version of the Human-Computer Trust Questionnaire developed by Madsen and Gregor (2000). Task shedding represented the number of times a participant shed tasks as well as the amount of time elapsed until they task shed. Performance on both the primary and secondary tasks was collected to ensure appropriate participant engagement in the tasks and to identify any outliers. Performance on the primary task was assessed as root-mean-square deviations from the tracking target. Performance on the secondary task was measured by time to agree or disagree (in secs) and appropriateness of agreement decision made. Rate of agreement with the automated decision aid was also measured. Finally, participant strategies and feedback were collected using an open-ended questionnaire created for this project (Appendix C).

Participants

An *a priori* power analysis using PASS 16 Power Analysis and Sample Size Software (2018) was completed based on $\alpha = 0.05$ to achieve a medium estimated effect size, Cohen's d = 0.5 (Cohen, 1992). This estimated effect was chosen based on effect sizes from similar research (partial $\eta^2 \geq .06$; Kaltenbach & Dolgov, 2017). Hancock et al. (2011) conducted a meta-analysis of factors affecting automation trust development and maintenance. For robot-related, performance-based factors (which include robot reliability and transparency), the researchers found a medium effect size from correlational studies ($r = .34$) as well as from experimental studies (Cohen's d = .71). An estimated N = 58 was needed to detect this effect size in a significant interaction. An α of 0.05 was chosen to balance the chance of Type I and Type II errors, to guard against interpreting a false effect as significant within a domain in which such errors could have serious consequences.

Participants were recruited from the undergraduate population at Old Dominion University through the Sona database. Participants were compensated with class credit for their participation. Participants were screened for normal or corrected-to-normal vision, and individuals with current or prior military experience were excluded from the participant pool to control for effects of military or task domain knowledge. Sixty-three participants were recruited for this study (47 female). Data for two participants were not used because one participant answered their phone during the study and the computer froze while running the other participant. Therefore 61 participants (45 female) were used for data analysis. Participant ages ranged from 18 to 28 (M = 20.02, SD = 2.15). There were 30 participants in the low reliability condition and 31 in the high reliability condition.

Materials

Demographic Form. Participant demographic information was collected to examine data to check for demographic effects. The questionnaire included age, sex, visual acuity and color blindness, vision correction (if applicable), computer use, handedness, and prior military experience (Appendix D).

Instruction Sheet. Participants received instructions regarding what to expect during the experiment as well as how to complete the MATB-II tracking task and the tank spotting task (Appendix E). The experimenter read the instructions aloud and the participant received a written copy to follow as well.

Vignettes. A vignette describing the ISR task domain and the observed reliability of the automated decision aid was provided and read aloud to each participant (Appendix B). The vignette provided background information concerning why the participant would be completing the experimental tasks. The vignette also included an explanation of system transparency as well as specific instructions for completing the tasks.

Trust Questionnaire. An adapted version of Madsen and Gregor's (2000) Human-Computer Trust (HCT) questionnaire was used to measure participants' subjective trust in the automated decision aid (Appendix F). For the six items chosen for this study, the wording of questions was adapted to use "tank spotting aid" rather than the original "system" wording, an adaptation similar to that used by Chancey (2016). To assess subjective trust, the HCT questionnaire provides the participant a statement such as, "I believe advice from the system even when I don't know for certain that it is correct," and the participant rates their agreement with the statement on a scale from 1 (*Not Descriptive*) to 12 (*Very Descriptive*).

The HCT questionnaire consists of 25 items, even divided across five dimensions: reliability, technical competence, understandability, faith, and personal attachment. The dimensions used in this study were reliability, understandability, and faith which reflected the dimensions of performance, process, and purpose described by Lee and See (2004), respectively. Reliability signifies the consistent, accurate functioning of the automated system. Understandability represents information that facilitates operator creation of a mental model of the automated system to predict future automated system behavior. Faith describes the operator's belief that automated system functioning will continue, even in situations that have not yet been encountered.

The six items used for the adapted HCT scale were chosen based on Madsen and Gregor's five-factor model by taking the two items with the highest loadings on their respective factors. From the five-factor analysis by Madsen and Gregor, the understandability (0.876 and 0.700) and faith (0.819 and 0.769) items showed good discriminant as well as convergent validity, loading only onto their respective factors. Reliability (0.628 and 0.533) items showed good convergent validity, but less discriminant validity by also loading onto the personal attachment (0.438 and 0.546) factor. The HCT questionnaire showed acceptable internal consistency. The overall questionnaire had a $\alpha_{Cronbach's}$ of 0.94, as well as subscale values for reliability ($\alpha_{Cronbach's}$ = 0.85), understandability ($\alpha_{Cronbach's}$ = 0.84), technical competence ($\alpha_{Cronbach's}$ = 0.74), faith ($\alpha_{Cronbach's}$ = 0.88), and personal attachment ($\alpha_{Cronbach's}$ = 0.90).

Following data collection from the present study, internal consistency was assessed using the collected data. The six-item scale showed good internal consistency ($\alpha_{Cronbach's}$ = .903). The two questions were highly correlated for reliability (r = .826, $p < .001$), understandability (r = .801, $p < .001$), and faith (r = .892, $p < .001$).

Multi-Attribute Task Battery II. The primary compensatory tracking task was programmed using the MATB-II and was presented on a desktop computer. MATB-II is a set of programmable tasks designed to simulate an aircraft cockpit during flight (Santiago-Espada et al., 2011). The tracking task required that the operator maintain a reticle within a target area by manipulating a joystick (see Appendix G for a tracking task image). The screen has a horizontal and a vertical bar, at the center of which is a target box. The participant had to keep the randomly drifting reticle at the center of the target box. Performance on the task was recorded as root-mean-square deviations from the center of the target, measured in pixels.

Tank-spotting Task. The secondary task was programmed using SuperEdit 4.7 software and presented through SuperCard 4.7 on a desktop Macintosh computer controlled with a standard mouse. The terrain images used were adapted from Chancey (2016; Appendix H). The task required that participants search a terrain image for a tank which may or may not be present, decide about tank presence or absence, or task shed the decision to automation. For all trials, participants had an automated decision aid which varied in information transparency and reliability. The automation provided an assessment of Tank Present or Tank Absent and reasoning for that assessment based on vehicle and traffic characteristics (see Appendix A for all transparency displays). For example, the transparency information for a Tank Present trial in low transparency displayed:

- Tank Present
- Conducted analysis of traffic patterns
- Traffic patterns are similar to those identified as hostile movement patterns

Each trial began with a blank screen for 3 seconds followed by the tank spotting task for 15 seconds. The task interface included the terrain image and automated decision aid transparency information, as well as Tank Present, Tank Absent, and Delegate Task buttons (see Appendix H for example of tank spotting image). The participant was able to make a tank presence decision at any point. If the participant chose tank present or tank absent, they were given feedback about the accuracy of the decision. The image remained on screen until the end of the 15 second trial before starting the next trial with the same procedure. If the participant chose to task shed, the automation decided in agreement with the transparency information provided. Accuracy feedback was immediately provided regarding the automation’s decision. Importantly, once the participant decided to task shed, the automation continued to make decisions until the end of the five-trial block (see Figure 4). The participant regained decision control at the beginning of the next block.

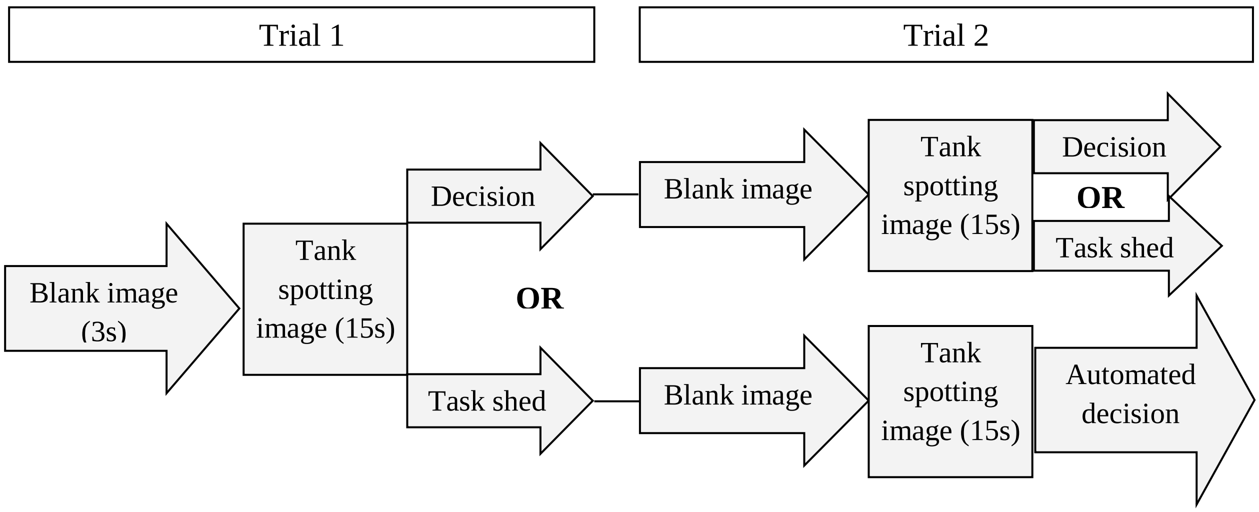


Figure 4. Progression of experimental trials. For example, in Trial 1 the participant saw a blank image followed by the tank spotting image and the options to indicate tank presence or task shed to the automated decision aid. If the participant decided, they would then follow the same procedure for Trial 2. If the participant task shed, the automation made the tank presence judgement for Trial 1 as well as the remaining trials in the block.

Post-Study Questionnaire. A brief questionnaire was given to participants after they finished all trials (Appendix C). The items included open-ended questions such as, “Did you use any specific strategies to complete the tank spotting task?” These were intended to elicit qualitative data to help explain participant reactions. The questionnaire also included a multiple-choice transparency manipulation check “Which line of additional information did NOT appear during the study?”

Apparati. The tracking task in MATB-II was presented on a desktop computer running the Windows 7 operating system. Participants controlled the tracking task using a Logitech Extreme 3D Pro joystick. The tank spotting task was run on a Macintosh desktop computer using OS X Yosemite Version 10.10.5 and presented on a 20-inch display. Participant task decisions were input with an Apple USB mouse (see Appendix I for images of experimental setup).

Procedure

After entering the laboratory, the participant was seated at a desk to first complete the Informed Consent Form (Appendix J) and a demographics form (Appendix D). The researcher then gave written instructions to the participant while also reading them aloud (Appendix E). Next, the participant completed separate training sessions for the tank spotting and MATB-II tracking tasks before practicing both tasks together. Training took approximately 5 minutes. For the tracking task training, participants used a joystick to maintain a randomly drifting reticle within a target box for 1 minute. The tank spotting training consisted of two five-image blocks resembling those in the experimental session, and the participant had to determine whether a tank was present. The participant would click “Tank Present” or “Tank Absent” and then receive feedback about the appropriateness of the response. If the participant had not task shed yet, the researcher instructed them to task shed on the third trial of the second block. This ensured that the participant understood how the task shedding process worked and allowed the researcher to point out the tank, ensuring that participants knew what the tanks looked like. Participants then trained on both tasks simultaneously. Next, the researcher read a vignette (Appendix B) which described the participant’s role and historical reliability of the automation while the participant read a written copy of the vignette. After reading the vignette and retrieving the written copy from the participant, the researcher asked the reliability manipulation check question, “Past performance has shown that this automation makes correct recommendations what percentage of the time?"

The experiment was organized into blocks with each block representing a single transparency condition. There were 4 blocks for each transparency condition (no, low, and high) for a total of 12 blocks. The order of presentation of transparency conditions was randomized to control for order effects. Each block comprised 5 trials.

During all 5 trials, the participant had to continuously monitor and control the tracking task. The tracking task and the first tank spotting trial were started at the same time, and the block ended after the participant completed the fifth tank spotting trial. The participant then completed the HCT questionnaire (Appendix F). Then the next block began. After all 12 blocks, participants were given a post-study questionnaire (Appendix C) and were debriefed before being dismissed.

In total, the experimental session lasted approximately 50 minutes. There was no concern for fatigue impacting participants' perception of transparency information due to the multi-line differences in transparency (0 lines, 2 lines, or 4 lines), which is a larger difference than typical one-line differences in much transparency research. A similar experimental task with similar time length has also been useful in automation trust research (Chancey, Bliss, Yamani, & Handley, 2017).

CHAPTER 3

RESULTS

Before analysis, data were assessed for normality by checking skewness (-1 to 1) and kurtosis (-2 to 2; Maxwell & Delaney, 2004). Five of the DVs violated the assumption of normality: age (skewness = 1.469, kurtosis = 2.584); average hours of daily computer use (skewness = 1.303, kurtosis = 2.665); and percentage of time task shed for none (skewness = 1.887, kurtosis = 3.099), low (skewness = 2.130, kurtosis = 3.870), and high transparency (skewness = 2.776, kurtosis = 8.566). Because many participants chose to retain active control, nearly any task shedding would result in non-normal data. However, the chosen analysis of variance (ANOVA) is generally robust to violations of normality (Maxwell & Delaney, 2004). Boxplots were used to identify outliers as values that were above or below the median by 1.5 interquartile range. Analysis was done with and without outliers (Mertler & Vannatta Reinhart, 2016). No difference in significance was found when outliers were removed; however, both values will be reported.

Manipulation check questions were examined. Participants correctly answered the reliability manipulation check 91.5% of the time, and the transparency manipulation check 73.8% of the time. Analyses were done with and without data from participants who failed the manipulation check questions. No differences in significance were observed.

Trust

Model assumptions were checked to ensure data were appropriate for parametric analyses. Levene's test demonstrated that data met the assumption of homogeneity of variance for all measures except time to task shed in high transparency, $F(1, 4) = 10.730$, $p = .031$. The assumption of sphericity was met by a non-significant Mauchly's test for all main analyses.

To evaluate the first hypothesis, a 2 × 3 mixed ANOVA was performed to assess the effects of transparency and reliability on operator trust. The ANOVA revealed a significant main effect of reliability on trust, $F(1, 59) = 36.622$, $p < .001$, partial $\eta^2 = .383$, observed power = 1.000, in which high reliability ($M = 9.49$, $SD = 2.45$) evoked higher trust than low reliability ($M = 6.92$, $SD = 3.12$; see Figure 5). There was no significant main effect of transparency, $F(2, 118) = .537$, $p = .586$, partial $\eta^2 = .009$, observed power = .137 or interaction, $F(2, 118) = 2.235$, $p = .112$, partial $\eta^2 = .036$, observed power = .448. This analysis was replicated with outliers removed, and the pattern of results was the same.

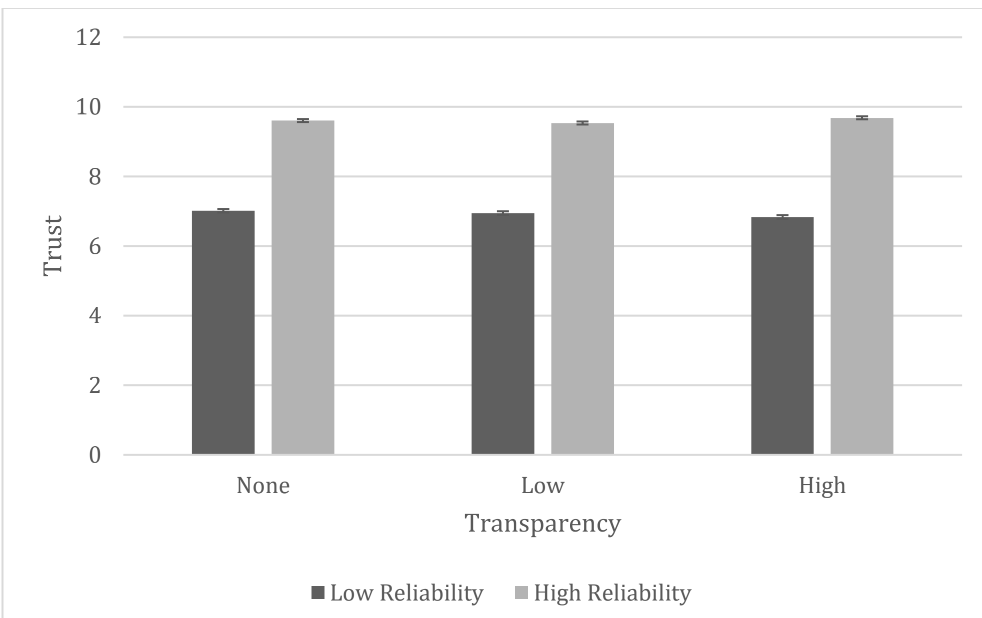


Figure 5. Effect of reliability on trust at each level of transparency. Error bars reflect standard error.

Operator trust has been demonstrated to calibrate over time (Lee & Moray, 1992; Merlo, 1999). To examine calibration, a repeated measures ANOVA was completed to assess the effects of time and reliability on operator trust. Mauchly's test showed a violation of the assumption of sphericity, $\chi^2 = 448.790$, $p = < .001$, Greenhouse-Geisser = .327. Using the Greenhouse-Geisser correction, there was a significant effect of time, $F(3.593, 197.633) = 9.391$, $p < .001$, partial $\eta^2 = .146$, observed power = .999 and a significant interaction of time and reliability, $F(3.593, 197.633) = 7.212$, $p < .001$, partial $\eta^2 = .116$, observed power = .992. This effect emerged as a significant linear trend of time, $F(1, 55) = 23.206$, $p < .001$, partial $\eta^2 = .297$, observed power = .997 and a significant linear trend of the time and reliability interaction, $F(1, 55) = 17.264$, $p < .001$, partial $\eta^2 = .239$, and observed power = .983 (see Figure 6).

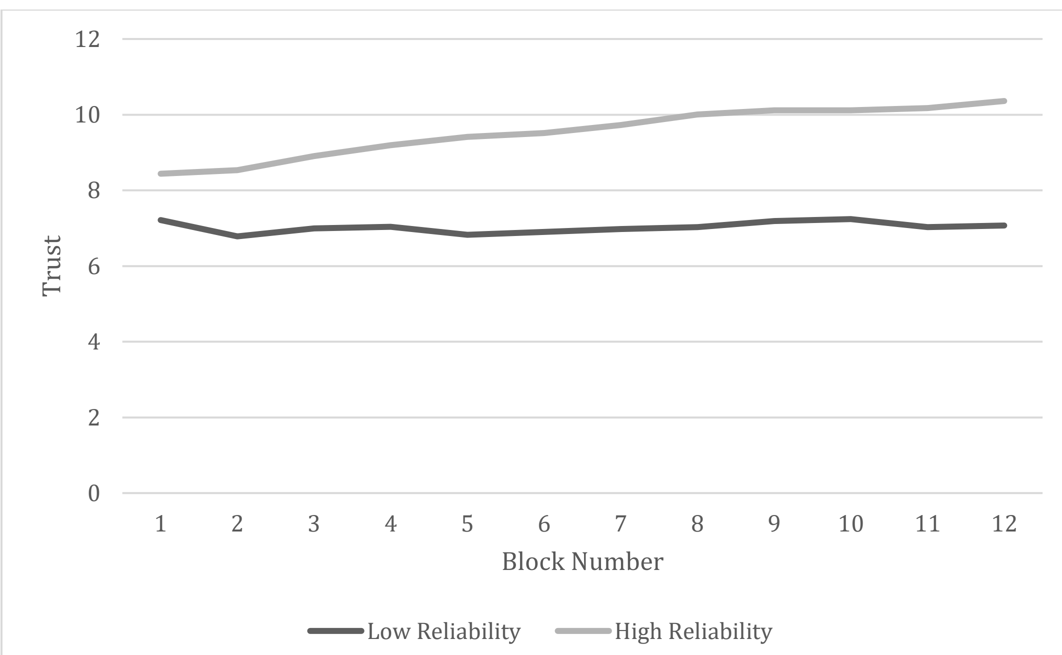


Figure 6. Graph of trust over time under high and low reliability conditions.

Because trust was demonstrated to calibrate across time, a repeated measures ANOVA was done on the trust measure from the last block in which the transparency condition was the same in both low and high reliability conditions. This calibrated trust may reflect more accurate participant trust compared to trust averaged across time. Examining a single trust measure this way is similar to studies that use a single trust measure at the end of the study (Bliss et al., 2013; Ma & Kaber, 2007). For low and high reliability, transparency conditions coincided for no transparency at block 9, for low at block 11, and for high at block 7 (see Table 1).

Table 1

Transparency Manipulations across Experimental Blocks for Low and High Reliability Groups

Block number	Low reliability	High reliability
1	None	Low
2	High	None
3	Low	Low
4	High	High
5	High	High
6	Low	None
7	High*	High*
8	None	None
9	None*	None*
10	Low	High
11	Low*	Low*
12	None	Low

Note: Asterisks indicate the latest block in which transparency aligned for low and high reliability groups.

A repeated measures ANOVA was performed, using calibrated trust in blocks 9, 11, and 7 to represent transparency levels none, low, and high, respectively. The data violated Mauchly's test of sphericity with Mauchly's $\chi^2 = 11.224$, $p < .004$, Greenhouse-Geisser = .848. The data met the homogeneity of variance assumption, $p > .05$. The main effect of transparency approached significance, $F(1.697, 98.411) = 3.059$, $p = .060$, partial $\eta^2 = .050$, observed power = .533; however, the interaction was not significant, $F(1.697, 98.411) = 1.652$, $p = .200$, partial $\eta^2 = .028$, observed power = .314. There was a significant linear trend of transparency, $F(1, 58) = 4.692$, $p = .034$, partial $\eta^2 = .075$, observed power = .568, in which increasing transparency resulted in decreasing trust (see Figure 7). The quadratic trend of the interaction approached significance, $F(1, 58) = 3.177$, $p = .080$, partial $\eta^2 = .052$, observed power = .418.

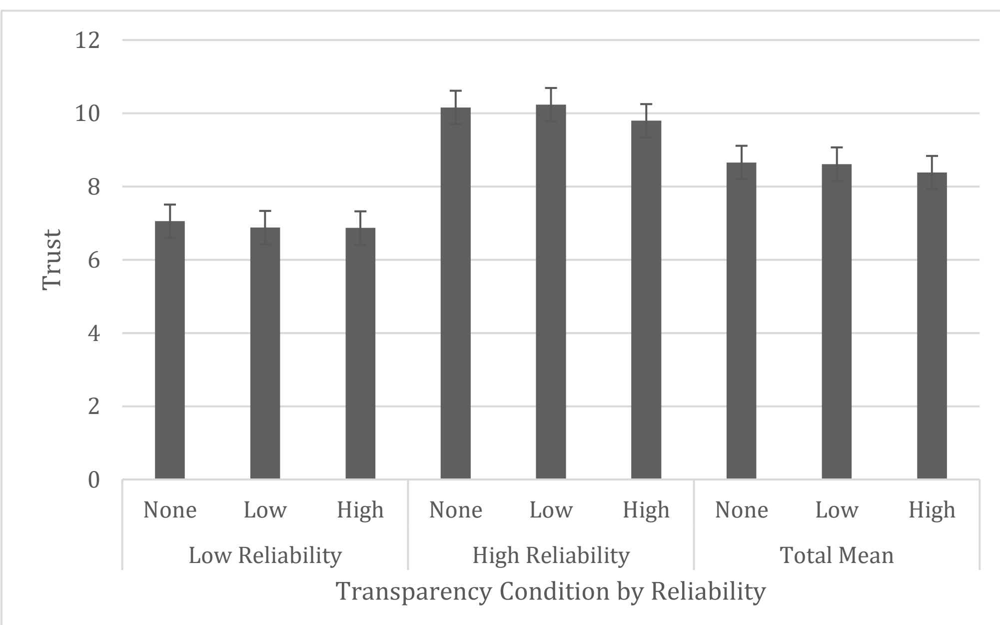


Figure 7. Effects of transparency and reliability on operator trust. Error bars reflect standard error.

Frequency of Task Shedding

To assess hypotheses two, three, and four, a repeated measures ANOVA was performed on the frequency of task shedding. The analysis revealed no significant main effect of transparency, $F(2, 118) = .409$, $p = .654$, partial $\eta^2 = .007$, observed power = .115; or reliability, $F(1, 59) = .009$, $p = .925$, partial $\eta^2 < .001$, observed power = .051; or interaction, $F(2, 118) = .409$, $p = .665$, partial $\eta^2 = .007$, observed power = .115.

This analysis was replicated without outliers. Data violated the homogeneity assumption at low ($F(1, 49) = 9.205$, $p = .004$) and high transparency ($F(1, 49) = 44.196$, $p < .001$). There was no significant effect of transparency ($F(2, 98) = .853$, $p = .429$, partial $\eta^2 = .017$, observed power = .193) or interaction ($F(2, 98) = 1.376$, $p = .258$, partial $\eta^2 = .027$, observed power = .290). The main effect of reliability approached significance, $F(1, 49) = 3.169$, $p = .081$, partial $\eta^2 = .061$, observed power = .415, demonstrated as higher frequency of task shedding in high reliability ($M = .07$, $SD = .13$) than low ($M = .03$, $SD = .07$).

Analysis was also done after arcsine transforming frequency of task shedding data. No difference in significance was observed.

Time to Task Shed

To assess hypothesis five, a univariate ANOVA was done to assess the effects of reliability on time to task shed. The analysis revealed no significant effect of reliability, $F(1, 24) = .157$, $p = .696$, partial $\eta^2 = .006$, observed power = .067. Analysis was also done after time data were log transformed, but showed no difference in significance.

Exploratory Analyses

Non-hypothesized, *post hoc* exploratory analyses were done to assess demographic effects on the collected data. A *t* test revealed that female participants reported less trust in automation, $t(59) = 2.523$, $p = .014$. Female participants also demonstrated worse flight tracking performance, $t(59) = -2.014$, $p < .049$. The effect of sex on tank spotting performance was approaching significance, $t(59) = 1.761$, $p = .083$. See Table 2 for all *t* test results, and Table 3 for descriptive statistics.

A multiple regression analysis was conducted to test whether age and average hours of daily computer use predicted frequency of task shedding. Similar analyses were conducted for time to task shed, trust, flight tracking performance, tank spotting performance, and tank spotting reaction time (RT). Age and computer use significantly explained 22% of the variance in tank spotting accuracy, $R^2 = .220$, $F(2, 58) = 8.162$, $p = .001$. Computer use significantly predicted tank spotting accuracy, $\beta = -.470$, $p < .001$. The predictive value of age on flight tracking performance approached significance, $\beta = .250$, $p = .055$. See Table 4 for full regression results.

Table 2

t Tests for Effect of Sex

				95% Confidence Interval	
Source	*t*	*df*	*p*	Lower Bound	Upper Bound
Trust	2.523	59	.014	.317	2.751
Task shed frequency	.193	24	.849	-2.473	2.984
TS	-1.072	59	.288	-.167	.050
MATB-II performance	-2.014	59	.049	-13.442	-.043
Tank spotting performance	1.761	59	.083	-.009	.135
Tank spotting reaction time*	.913	59	.365	-.605	1.621

Table 3

Means and Standard Deviations for Dependent Variables

	Sex	*n*	*M*	*SD*
Trust	Male	16	9.42	1.59
	Female	45	7.89	2.23
Time to task shed	Male	6	7.40	2.90
	Female	20	7.14	2.82
Task shed frequency	Male	16	.07	.10
	Female	45	.13	.21
Flight tracking deviations from center	Male	16	42.98	11.76
	Female	45	49.72	11.42
Tank spotting performance	Male	16	.83	.10
	Female	45	.77	.13
Tank spotting reaction time*	Male	16	8.16	1.51
	Female	45	7.65	2.03

*Levene's $F = 4.44$, $p = .04$

Trust and agreement with the automated decision aid were significantly correlated, $r = .595$, $p < .001$. However, task shedding frequency was not significantly correlated with trust ($r = .143$, $p = .271$) or agreement with automation ($r = .214$, $p = .097$).

Table 4

Regression Results for Effects of Age and Average Daily Computer Use

		t	*p*	*β*	*F*	*df*	*p*	R^2
Frequency of task shedding					.214	2, 58	.808	.007
	Age	-.037	.971	-.005				
	Computer Use	.649	.519	.085				
Time to task shed					.357	2, 23	.704	.030
	Age	.498	.623	.102				
	Computer Use	-.670	.509	-.138				
Trust					.145	2, 58	.866	.005
	Age	.413	.681	.054				
	Computer Use	-.312	.756	-.041				
Flight tracking performance					2.021	2, 58	.142	.065
	Age	1.960	.055	.250				
	Computer Use	-.295	.759	-.038				
Tank spotting performance					8.162	2, 58	.001	.220
	Age	-.337	.737	-.039				
	Computer Use	-4.040	.000	-.470				
Tank spotting RT					2.381	2, 58	.101	.076
	Age	1.176	.244	.149				
	Computer Use	-1.740	.087	-.220				

CHAPTER 4

DISCUSSION

The goal of this study was to examine how transparency and reliability interact to influence task shedding behavior and operator trust in automation. By controlling the amount of transparency, designers may be able to facilitate operators as they demonstrate appropriate trust and automation use to improve human-automation system performance. Transparency may be used across different reliability levels to inform operators' decisions to task shed to minimize the danger of automation use errors.

Trust

Hypothesis one predicted an interaction of transparency and reliability on operator trust, as manifested by self-report data. This hypothesis was not supported. Transparency did not significantly affect trust in either the high or low reliability conditions. Supporting previous research, high reliability did increase trust (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Hancock et al., 2011; Helldin et al., 2013; Ma & Kaber, 2007; Ososky et al., 2014; Parasuraman & Riley, 1997). There was no effect of transparency (singly or jointly with reliability) on trust.

The lack of effect of transparency on subjective trust could have resulted from two major visual attention demands on the participant. Because the flight tracking task required frequent monitoring, participant attention may have been dominantly focused on the primary task as well as on the tank spotting image, leaving little attention to read the transparency information beyond the automation decision recommendation. Additionally, the short time on each tank spotting trial may have limited participants' ability to process the transparency information.

Participant feedback indicated an influence of time and attention demands. One participant reported scanning each quadrant of a tank spotting image for about three seconds, leaving only three seconds to attend to both the flight tracking task and transparency information. Other participants explained that the tasks were overwhelming or that it was difficult to pay attention to both, such as "In the moment I trusted the [automation] more because I wasn't able to give my full attention." Another stated, "I would count in my head to 15 seconds to make sure I didn't run out of time."

Further examination of calibrated trust did reveal a linear trend of declining trust as transparency increased. This may have been a result of the amount of information that was wrong when errors occurred. When automation errors occurred in high transparency, participants had more false information, compared to less incorrect information in lower transparency conditions.

These findings reflect the body of research demonstrating the mixed influence of transparency on trust. Past research has shown an effect of transparency when operators are given graphical sensor data along with text-based transparency and unit category recommendation; although no effect was shown with text only or with text and class recommendation information (Helldin, 2014). Other research failed to find an effect (Chen et al., 2015) or found a variable effect depending on the trust measurement used (Kaltenbach & Dolgov, 2017).

Although the effect of transparency varied with trust calibration, two participants stated that the amount of transparency directly affected their opinion of the automation: "At first it seemed as though the shortest descriptions were the most accurate" and "I was more likely to agree with it when it gave more information." This may indicate biases in operator reactions to transparency information that should be further studied.

Frequency of Task Shedding

Hypotheses two, three, and four predicted that transparency and reliability would increase the frequency of task shedding. These hypotheses were not supported. Contrary to expectations and pilot study observations, participants generally demonstrated an unwillingness to relinquish task control and a willingness to accept automation recommendations *prima facie*. This supports research that has demonstrated an operator bias to retain personal control (Parasuraman & Riley, 1997). Participant feedback supported this bias, such as "I never used it [task shedding] because it felt like giving up" and "I refused to [task shed], if lives are at risk a computer with 10% chance to fail kills 10% of the people you want to protect."

Self-confidence also increases operator retention of task control (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Parasuraman & Riley, 1997). In this study, participants may have felt sufficiently confident in their tank spotting accuracy, reducing their likelihood to rely on the automation. Some participants chose to retain control at the cost of time-out errors. Twenty-two participants had at least one time-out error, and fourteen had more than one.

Similarly, participants may have used the transparency information without task shedding. In the post-study questionnaire, 32 participants reported using the automation recommendation (e.g. "Though I didn't use the delegate option, I still based my responses on the analysis."). Because of the visual demand of monitoring both the primary tracking task and the tank spotting image, participants may have read only the recommendation but not the supporting transparency information. Research by Wright and colleagues (2017) used eye tracking to examine how participants attend to transparency information. Future research could include eye tracking to study transparency in visually demanding situations.

Time to Task Shed

Hypothesis five predicted that increased reliability would result in faster task shedding. This hypothesis was not supported. This may be due to participants' hesitance to task shed in general. If a participant tends to not task shed at all, time taken to task shed may lose sensitivity as a dependent measure.

This finding particularly contributes to research examining the relationship between trust and task shedding. Although participants were more trusting of highly reliable automation, trust did not translate to increased automation use in this situation. Trust has been demonstrated by other researchers to be a precursor to task shedding (cf., Bliss, Harden, & Dischinger, 2014), but this relationship did not hold in the current experiment. Other factors such as self-confidence or use of decision recommendation without fully relinquishing task control could have contributed to these findings. McGuirl and Sarter (2006) found that participants would use an automated decision support system as a warning but did not rely on it for a final decision. Self-confidence influences automation use, possibly reducing the effect of trust on task shedding in this study (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Parasuraman & Riley, 1997). A few participants specifically stated they thought they were more accurate than the automation or that they had confidence in their own abilities.

Theoretical Implications

Although research conducted by Barnes et al. (2017), Helldin et al. (2013), and Ososky et al. (2014) has found effects of transparency on trust, this was not demonstrated in the current experiment. Notably, other research has shown mixed results with an effect at only high transparency (Helldin, 2014), a positive relationship with only some dimensions of trust (Chen et al., 2015), or an effect dependent on trust scale used (Kaltenbach & Dolgov, 2017). Such findings reflect the complex nature of the constructs of transparency and trust.

The finding that high reliability, compared to low, increased operator trust in automation supports the general conclusions of human-automation research (Chavaillaz et al., 2016; de Visser & Parasuraman, 2011; Hancock et al., 2011; Helldin et al., 2013; Ma & Kaber, 2007; Ososky et al., 2014; Parasuraman & Riley, 1997). This study also demonstrated that reliability influences how operator trust calibrates over time. Although initial trust may be similar for participants in low and high reliability groups, over time trust will increase as participants interact with more reliable automation. The lack of correlation between trust and task shedding also demonstrates that self-reported trust is only one predictor of operator behavior.

Automation bias is the operator tendency to use automation without calibrated trust guiding automation use (McGuirl & Sarter, 2006). One example of this was participants who immediately relinquish a task to automation (Bliss et al., 2013), a behavior not generally demonstrated in this study. A second bias in automation use is operator retention of task control (Parasuraman & Riley, 1997). This tendency was evident here regardless of reliability and transparency conditions, occurring even at the cost of timeout errors.

According to Lyons and Havig (2014), transparency should improve an operator's mental model of automation. Were that the case here, participants would have performed better with more transparency and would task shed to highly reliable automation. In contrast, transparency level did not affect performance or task shedding. It follows that transparency may not have effectively influenced development of mental models. Lyons (2013) emphasizes the importance of training for operators to understand transparency in the intentional and analytical models.

Such training may influence operators' ability to utilize transparency when forming accurate mental models of automation. The findings of this study demonstrated that novice participants did not effectively incorporate transparency into their automation use decisions.

Practical Implications

Broadly, these findings benefit practical applications by demonstrating that the effects of transparency on task shedding and trust may be influenced by the specific situations in which operators interact with automation. Any effects of transparency on operator behavior may be masked in applied tasks that feature high attention demand or workload, such as air traffic control, nuclear power operation, or aircraft piloting. In such cases, examining the salience of transparency may be vital to ensure that operators are attending to the information.

In the specific realm of military ISR, automation has been proposed to improve system performance by providing fast data selection and analysis, assimilation of data sources, and action recommendations (Adams et al., 2003; Parasuraman et al., 2000; Tyler, 1999). However, transparency may be difficult to implement in ISR tasks that require continuous monitoring of surroundings. In these cases, transparency presentation in modalities other than text may be beneficial for operator attention (Kilgore & Voshell, 2014; Sanders et al., 2014). Another possibility is to utilize likelihood alarm signals to embed reliability information within discrete indicators (Sorkin, Kantowitz, & Kantowitz, 1988).

The findings of the current research demonstrated that operators calibrate trust over time, indicating that operators should spend time interacting with automation before making the choice to use automation or not. This may be particularly beneficial to reduce disuse of highly reliable automation. However, trust is only one factor impacting automation use decisions and should be considered along with other influences to encourage task shedding to highly reliable automation or to discourage use of unreliable automation.

Limitations

One limitation in this study was balancing the difficulty of the primary task. Though task difficulty is necessary to evoke task shedding, some participants may have been overwhelmed and unable to process the transparency information. This limitation may be circumvented in future studies by presenting the primary task and transparency information in different sensory modalities or by retaining a high attention task while reducing the visual workload.

Another limitation to be addressed was the absence of a temporal progress bar which was left out for technical reasons. The lack of a progress bar may have introduced uncertainty. Uncertainty could increase workload or increase the likelihood participants would rely on individual biases such as overconfidence or misrepresentation of error rates while completing the tank spotting task (Parasuraman & Riley, 1997; Tversky & Kahneman, 1974).

To incorporate misses as well as false alarms, training blocks were 60% reliable. This may have influenced trust calibration during the early blocks of the experiment. However, participants were told the automation reliability rate, a factor that has facilitated trust calibration in past research (Bliss, 1993) but that could mask the effect of 60% reliable training. Also regarding reliability, due to the number of trials per block, half of the high reliability blocks contained one error and half contained zero errors, which averaged to 90% reliable across all blocks. In this condition, participants experienced changing reliability levels, a factor that may influence trust (Wiegmann et al., 2001).

Future Directions

Based on the experimental results reported here, two major directions of research emerge: the relationship between transparency and cognitive workload and how task situation may influence automation use behaviors. Factors like task criticality and the need to perform multiple tasks concurrently can influence operator trust and automation use (Parasuraman & Riley, 1997; Wickens et al., 2006). These factors should be examined with task shedding as a dependent measure to more clearly identify situations in which operators are willing to relinquish task control.

The relationship between transparency and cognitive workload could be complex, perhaps mediated by information utility. Evidence for an effect of transparency on operator workload is mixed. Theoretical explorations of transparency have predicted that situations that feature greater transparency will increase workload (Lyons, 2013; Ososky et al., 2014). Chen and colleagues (2015) found that increasing transparency resulted in an increase in the mental demand and frustration dimensions of the NASA Task Load Index (NASA-TLX). Conversely, some studies have not found effects of the content or modality of transparency on workload (Barnes et al., 2017; Chen et al., 2017; Sanders, Wixon, Schafer, Chen, & Hancock, 2014; Selkowitz, Lakhmani, Larios, & Chen, 2016). One explanation for the effect of transparency on workload is that increasing transparency may decrease workload by reducing situation uncertainty, thereby making the operator's decision easier. However, increases in transparency could result in a greater amount of information to process, thereby increasing workload.

Workload may also influence the usefulness or effectiveness of transparency. High demand may reduce the operator's ability to attend to transparency information. Because of this, research should examine how workload and transparency interact relative to task and transparency modalities. The transparency manipulation may influence operators' ability to process the information. For high visual demand tasks, auditory or pictorial information may communicate transparency better than text. Ultimately (particularly in light of the current findings), more research is needed to understand the underlying relationship between transparency and workload.

A final consideration in automation use is neglect tolerance, or the amount of time an unmanned autonomous entity can function unaided before performing below a given threshold (Olsen & Goodrich, 2003). Because automation is often imperfect, neglect tolerance could be one factor influencing automation use behaviors like task shedding. It could be that operators would not task shed to automation with short neglect tolerance because they would need to resume control sooner. However, such disuse could negatively impact human-automation performance if the automation is highly reliable. Transparency information could help the operator calibrate task shedding relative to neglect tolerance or guide the operator to intervene sooner as automation performance decreases.

Conclusions

Automation use can have a large influence on the performance of human-automation systems. The goal of this study was to understand how transparency of automated processes and reliability of automation influence operator trust and task shedding. Following from previous researchers, analyses demonstrated successful manipulation of self-report trust by advertisement of information reliability. Results concerning the role of information transparency, however, were mixed. This may underscore the complex relationship among transparency, reliability, trust, and related constructs. Although transparency may be beneficial, the degree of benefit may vary across situations. Future research is needed to fully understand how designers can contribute to beneficial human-automation system performance.

REFERENCES

Adams, B. D., Bruyn, L. E., Houde, S., & Angelopoulos, P. (2003). *Trust in automated systems literature review.* Defense Research and Development Canada, Toronto.

Barnes, M. J., Chen, J. Y., & Hill, S. (2017). *Humans and autonomy: Implications of shared decision making for military operations* (Report No. ARL-TR-7919). Retrieved from Defense Technical Information Center website: https://apps.dtic.mil/docs/citations/AD1024840

Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 24–29). Hillsdale, NJ: Lawrence Erlbaum.

Bliss, J.P. (1993). *The cry-wolf phenomenon and its effect on alarm response* (Unpublished doctoral dissertation). University of Central Florida, Orlando, FL.

Bliss, J. P., & Gilson, R. D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics, 41*, 57-72.

Bliss, J. P., Harden, J. W., & Dischinger Jr, H. C. (2013). Task shedding and control performance as a function of perceived automation reliability and time pressure. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*, 635-639. Los Angeles, CA: Sage. doi:10.1177/1541931213571136

Boubin, J. G., Rusnock, C. F., & Bindewald, J. M. (2017). Quantifying compliance and reliance trust behaviors to influence trust in human-automation teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*, 750-754. Los Angeles, CA: Sage. doi:10.1177/1541931213601672

Chancey, E. T. (2016). *The effects of alarm system errors on dependence: Moderated* *mediation of trust with and without risk* (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

Chancey, E. T., Bliss, J. P., Proaps, A. B., & Madhavan, P. (2015). The role of trust as a mediator between system characteristics and response behaviors. *Human Factors, 57,* 947-958. doi:10.1177/0018720815582261

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, 59*, 333-345. doi:10.1177/0018720816682648

Chancey, E.T., Proaps, A., & Bliss, J.P. (2013). The role of trust as a mediator between signaling system reliability and response behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*, 285-289. Los Angeles, CA: Sage. doi:10.1177/1541931213571063

Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics, 52,* 333-342. doi:10.1016/j.apergo.2015.07.012

Chen, J. Y., Barnes, M. J., Wright, J. L., Stowers, K., & Lakhmani, S. G. (2017). Situation awareness-based agent transparency for human-autonomy teaming effectiveness. *Proceedings Volume 10194, Micro-and Nanotechnology Sensors, Systems, and Applications IX, 10194*. Anaheim, CA: International Society for Optics and Photonics.doi:10.1117/12.2263194

Chen, T., Campbell, D., Gonzalez, L. F., & Coppin, G. (2015). Increasing autonomy transparency through capability communication in multiple heterogeneous UAV management. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2434-2439. Hamburg, Germany: IEEE.doi:10.1109/IROS.2015.7353707

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (Report No. ARL-TR-6905). US Army Research Laboratory Aberdeen Proving Ground United States.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155. doi:10.1037/0033-2909.112.1.155

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*, 209-231. doi:10.1177/1555343411410160

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65-84. doi:10.1518/001872095779049499

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human factors*, *59*, 5-27. doi:10.1177/0018720816681350

Foyle, D.C., Ahumada, A.J., Larimer, J. and Sweet, B.T. (1992). Enhanced/synthetic vision systems: Human factors research and implications for future systems. *SAE Transactions: Journal of Aerospace*, *101*, 1734-1741. doi:10.4271/921968

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*, 517-527. doi:10.1177/0018720811417254

Helldin T., Falkman G., Riveiro M., Dahlbom A., & Lebram M. (2013) Transparency of military threat evaluation through visualizing uncertainty and system rationale. *Proceedings of the 10th International Conference Engineering Psychology and Cognitive Economics, 8020,* 263-272. Heidelberg, Germany: Springer. doi:10.1007/978-3-642-39354-9_29

Hillesheim, A. J., & Rusnock, C. F. (2016). Predicting the effects of automation reliability rates on human-automation team performance. *Proceedings of the 2016 Winter Simulation Conference*, 1802-1813. Arlington, VA: IEEE Press.

Kaltenbach, E., & Dolgov, I. (2017). On the dual nature of transparency and reliability: Rethinking factors that shape trust in automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*, 308-312. Los Angeles, CA: Sage. doi:10.1177/1541931213601558

Kilgore, R. & Voshell, M. (2014). Increasing the transparency of unmanned systems: Applications of ecological interface design. *International Conference on Virtual, Augmented and Mixed Reality*, 378-389. Springer International Publishing. doi:10.1007/978-3-319-07464-1_35

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153-184. doi:10.1006/ijhc.1994.1007

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*, 50-80. doi:10.1518/hfes.46.1.50_30392

Livingston, M. A., Rosenblum, L. J., Julier, S. J., Brown, D., Baillot, Y., Swan, I. I., ... & Hix, D. (2002). An augmented reality system for military operations in urban terrain. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference, Orlando, FL,* 89-97.

Livingston, M. A., Swan, J. E., Gabbard, J. L., Hollerer, T. H., Hix, D., Julier, S. J., ... & Brown, D. (2003). Resolving multiple occluded layers in augmented reality. *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, 56-65. Tokyo, Japan: IEEE. doi:10.1109/ISMAR.2003.1240688

Lyons, J. B. (2013). Being transparent about transparency: A model for human–robot interaction. In D. Sofge, G. J. Kruijff, & W. F. Lawless (Eds.), *Trust and autonomous systems: Papers from the AAAI Spring Symposium* (Tech. Rep. SS-13-07). Menlo Park, CA: AAAI Press.

Lyons, J. B., & Havig, P. R. (2014. Transparency in a human-machine context: Approaches for fostering shared awareness/intent. *International Conference on Virtual, Augmented and Mixed Reality*, 181-190. Switzerland: Springer. doi:10.1007/978-3-319-07458-0_18

Ma, R., & Kaber, D. B. (2007). Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. *International Journal of Industrial Ergonomics*, *37*, 665-673. doi:10.1016/j.ergon.2007.04.005

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*, 277-301. doi:10.1080/14639220500337708

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. *Proceedings of the Eleventh Australasian Conference on Information Systems, 53*, 6-8. Brisbane, Australia: Australasian Association for Information Systems.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*, 123-136. doi:10.1037/0021-9010.84.1.123

Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*, 709-734. doi:10.2307/258792

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors, 48*, 656-665. doi:10.1518/001872006779166334

Merlo, J. L., Wickens, C., & Yeh, M. (2000). Effect of reliability on cue effectiveness and display signaling. *Proceedings of the 4th Annual Army Federated Laboratory Symposium*. College Park, MD: Army Research Laboratory.

Mertler, C. & Vannatta Reinhart, R. (2016). *Advanced and multivariate statistical methods: Practical application and interpretation* (6th ed.). New York, NY: Routledge.

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

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, *49*, 57-75. doi:10.1518/001872007779598037

Moray, N., & Inagaki, T. (1999). Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control*, *21*(4-5), 203-211. doi:10.1177/014233129902100408

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*, 44-58. doi:10.1037/1076-898X.6.1.44

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*(5-6), 527-539. doi:10.1016/S0020-7373(87)80013-5

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. doi:10.1080/00140139408964957

Olsen, D. R., & Goodrich, M. A. (2003). Metrics for evaluating human-robot interactions. *Proceedings of Performance Metrics for Intelligent Systems, Gaithersburg, MD.*

Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. Y. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *SPIE Defense+ Security*, 90840E-90840E. Bellingham, WA: International Society for Optics and Photonics. doi:10.1117/12.2050622

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems* (Report No. NAWCADWAR-92033-60) Washington, DC: The Catholic University of America.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (2005). A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, *35*, 481–493. doi:10.1109/TSMCA.2005.850598

Parasuraman, R. & Hancock, P. A. (2001). Adaptive control of mental workload. In B.H. Kantowitz (Eds.), *Stress, workload, and fatigue* (pp. 305-320). London, U.K.: Lawrence Erlbaum Associates, Inc.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. *Automation Technology and Human Performance: Current Research and Trends*, 119-123. doi:10.1177/001872088803000405

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*, 230-253. doi:10.1518/001872097778543886

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. doi:10.1109/3468.844354

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, *50*, 511-520. doi:10.1518/001872008X312198

PASS 16 Power Analysis and Sample Size Software (2018). [Computer Software]. Kaysville, UT: NCSS, LLC. ncss.com/software/pass.

Rice, S. (2009) Examining single- and multiple-process theories of trust in automation. *The Journal of General Psychology, 136*, 303-322. doi:10.3200/GENP.136.3.303-322

Rouse, W.B. (1994). Twenty years of adaptive aiding: Origins of the concept and lessons learned. In M. Mouloua and R. Parasuraman. Eds., *Human performance in automated systems: Current research and trends*. (pp. 28-32), Hillsdale, NJ, Erlbaum.

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, *49*, 76-87. doi:10.1518/001872007779598082

Sanders, T. L., Wixon, T., Schafer, K. E., Chen, J. Y., & Hancock, P. A. (2014). The influence of modality and transparency on trust in human-robot interaction. *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA),* 156-159. San Antonio, TX: IEEE. doi:10.1109/CogSIMA.2014.6816556

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).* Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.

Scerbo, M. W. (1994). Implementing adaptive automation in aviation: The pilot-cockpit team. *Human Performance in Automated Systems: Current Research and Trends*, 249-255.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. *Automation and Human Performance: Theory and Applications* (pp. 37-63). Hillsdale, NJ: Lawrence Erlbaum Associates.

Selkowitz, A. R., Lakhmani, S. G., Larios, C. N., & Chen, J. Y. (2016). Agent Transparency and the Autonomous Squad Member. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60,* 1319-1323. Los Angeles, CA: Sage doi:10.1177/1541931213601305

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators.* MIT Man-Machine Systems Laboratory Report.

Sorkin, R.D., Kantowtiz, B., & Kantowitz, S. (1988). Likelihood alarm displays. *Human Factors, 30*, 445-459. doi:10.1177/001872088803000406

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi:10.1126/science.185.4157.1124

Tyler, R. R. (1999). Human automation interaction - a military user's perspective. In M. W. Scerbo & M. Mouloua (Eds.) *Automation Technology and Human Performance: Current Research and Trends* (pp. 38-41). Mahwah, NJ: Lawrence Erlbaum Associates.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human machine systems II: Mediating “direct perception” in complex work domains. *Ecological Psychology*, 2, 207-249.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors*, *51*, 281-291. doi:10.1177/0018720809338842

Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team: Comparing automatically generated explanations. *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 109-116). Piscataway, NJ: IEEE Press.

Wickens, C. D., Dixon, S. R., & Ambinder, M. S. (2006). Workload and automation reliability in unmanned air vehicles. In N. Cooke, H. L. Pringle, H. K. Pedersen, & O. Connor (Eds.), *Human factors of remotely operated vehicles* (pp. 209-222). Amsterdam: Elsevier.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors, 2*, 99–126. doi:10.1207/STHF0202_01

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, *2*, 352-367. doi:10.1080/14639220110110306

Wright, J. L., Chen, J. Y., Barnes, M. J., & Hancock, P. A. (2017). The effect of agent reasoning transparency on complacent behavior: An analysis of eye movements and response performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*, 1594-1598. Los Angeles, CA: Sage. doi:10.1177/1541931213601762

APPENDIX A

DECISION AID TRANSPARENCY OUTPUT

No Transparency
Tank Present

OR

Tank Absent

Low Transparency
Tank Present
Conducted analysis of traffic patterns
Traffic patterns are similar to those identified as hostile movement patterns

OR

Tank Absent
Conducted analysis of traffic patterns
Traffic patterns are unlike to those identified as hostile movement patterns

High Transparency
Tank Present
Conducted analysis of metallic signatures
Conducted analysis of traffic patterns
Strength of metallic signature exceeds minimum requirement for identification
Traffic patterns are similar to those identified as hostile movement patterns

OR

Tank Absent
Conducted analysis of metallic signatures
Conducted analysis of traffic patterns
Strength of metallic signature does not meet minimum requirement for identification
Traffic patterns are unlike those identified as hostile movement patterns

APPENDIX B

VIGNETTES

For this experiment, you will assume the role of an Information, Surveillance, and Reconnaissance analyst. Insurgents in Kandahar, Afghanistan have been purchasing old Russian T-72 tanks. Your job as the analyst is to look through static satellite images of terrain, searching for potential targets. Along with the image, an automated decision aid will provide you with a "Tank Present" or "Tank Absent" recommendation. The automation may also provide information explaining why the recommendation was made. Past performance has shown that this automation makes correct recommendations 60% [or 90% in high reliability condition] of the time. Errors may consist of a false alarm indicating a tank is present when there is no tank, or a miss indicating there is no tank when there is a tank present. Your job is to make a decision whether there is a tank present or not. You may also delegate the task to the automation, in this case the automation will follow its recommendation. Due to the sensitive nature of this task, it is important that you make an accurate decision.

APPENDIX C

POST-STUDY QUESTIONNAIRE

Participant number:______________ Date:_____________

1. Did you have any specific strategies for completing the tracking task?

2. Did you have any specific strategies for completing the tank spotting task?

3. How did you feel about delegating the task to the automated decision aid?

4. Which line of additional information did NOT appear during the study?

    a. Traffic patterns are similar to those identified as hostile movement patterns

    b. Conducted analysis of vehicle weight

    c. Conducted analysis of traffic patterns

    d. Strength of metallic signature does not exceed minimum requirement for identification

5. Any other comments:

APPENDIX D

DEMOGRAPHIC FORM

Participant #: _______ Date: _____________

The purpose of this questionnaire is to collect background information. The information provided is strictly for the purposes of research only.

1. Age:________
2. Sex: Male Female Other
3. Which hand do you predominantly use? Right Left Ambidextrous
4. Have you ever been diagnosed as having a deficiency in your visual acuity (less than perfect vision)? Yes No

   If yes, do you have correction (i.e. glasses, contact lenses, etc.) with you?

   Yes No
5. Have you ever been diagnosed as color deficient or color blind? Yes No
6. Indicate the average number of hours per day you spend using computers (personal and work combined):___________
7. Do you have any prior military service? Yes No

   If yes, please explain:____

   ____

APPENDIX E

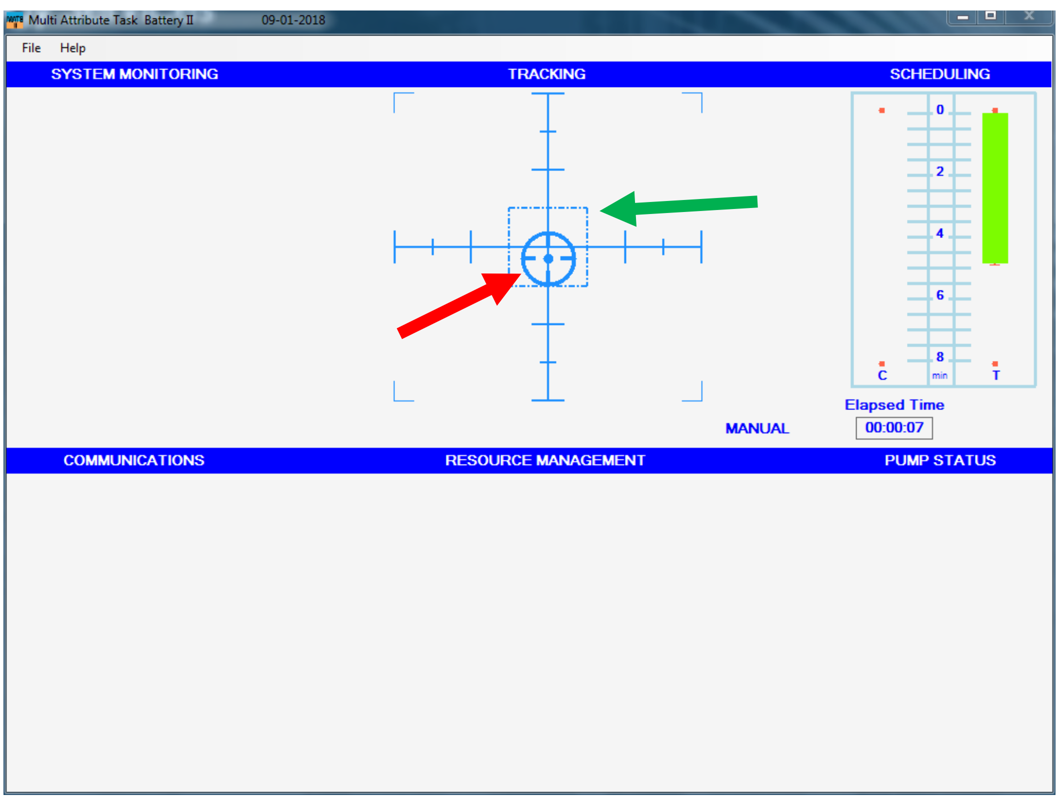
INSTRUCTION SHEET

Welcome to the REACTS Lab. Thank you for participating in the study today, it should take about 60 minutes and you will receive 1.5 Sona credits for your participation.
Please silence your cell phone and put it away for the duration of the study.

For this study, you will perform two tasks: a tank spotting task and a flight simulation tracking task. The tasks will be performed simultaneously, and your performance will be recorded for analysis by the researcher.

Primary Tracking Task

For the tracking task, you will use a joystick to control a continuously drifting target within a center box. This task is similar to a flight simulation task in which you guide an aircraft (drifting target) along a target path (center box). Please try to keep the target at the center of the box for the duration of the study. If the target leaves the box, use the joystick to move it back to the center of the box.



Above is an example of what the tracking task will look like. The target (red arrow) and box (green arrow) are indicated.
Do you have any questions?

Secondary Tank Spotting Task

For the tank spotting task, you will see a satellite image of terrain in a warzone. Your task is to identify whether or not the image contains an enemy tank with the help of an automated decision aid. You have 15 seconds to click either a “Tank” or a “No Tank” button or you may choose to delegate this decision to an automated decision aid by clicking a “Delegate” button. If you choose to delegate, the automated decision aid will continue to make decisions until the next set of images.

Above are some examples of what tanks may look like.

This is an example of what a terrain image containing a tank (circled in red) can look like.

Above is an example of what the interface will look like, in this case no tank is present. The blank box indicates where the automated decision aid will provide a decision recommendation and may provide additional information. Below is an example of additional information that the decision aid may present:

Conducted analysis of metallic signatures
Conducted analysis of traffic patterns
Strength of metallic signature exceeds minimum requirement for identification
Traffic patterns are similar to those identified as hostile movement patterns

Buttons are available to indicate whether there is or is not a tank present. There is also a button to delegate the decision to the automated decision aid.

Do you have any questions?

Next you will complete practice trials on the tank spotting and tracking tasks individually, followed by practice with both tasks simultaneously.

Experiment

For the duration of the experimental session, you will complete both tasks simultaneously. For each set of trials, you will continuously perform the primary tracking task while processing 5 tank spotting images.

As you do the tracking task, you will be presented with a tank spotting image, and will have 15 seconds to choose “Tank” or “No Tank” or to “Delegate” the decision to the automated decision aid. Once you make a decision or choose to delegate, you will receive feedback regarding the accuracy of your decision while the image will remain onscreen for the remainder of the 15 seconds. If you have delegated the decision to the automated decision aid, you will see feedback about the accuracy of the automation.

After the feedback, another tank spotting image will appear. If you delegated the previous task, the automation will complete this and any following trials. If you did not delegate, you will make a decision just like the first image. There are 5 images total.

After 5 images, you will answer a brief questionnaire. You will then start the next session of tracking and tank spotting tasks.

Do you have any questions before you begin?

APPENDIX F

MADSEN AND GREGOR TRUST QUESTIONNAIRE (2000)

2. Perceived Reliability

R1 - The system always provides the advice I require to make my decision.
*R2 - The system performs reliably.
R3 - The system responds the same way under the same conditions at different times.
*R4 - I can rely on the system to function properly.
R5 - The system analyzes problems consistently.

3. Perceived Technical Competence

T1 - The system uses appropriate methods to reach decisions.
T2 - The system has sound knowledge about this type of problem built into it.
T3 - The advice the system produces is as good as that which a highly competent person could produce.
T4 - The system correctly uses the information I enter.
T5 - The system makes use of all the knowledge and information available to it to produce its solution to the problem.

4. Perceived Understandability

U1 - I know what will happen the next time I use the system because I understand how it behaves.
*U2 - I understand how the system will assist me with decisions I have to make.
*U3 - Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.
U4 - It is easy to follow what the system does.
U5 - I recognize what I should do to get the advice I need from the system the next time I use it.

5. Faith

*F1 - I believe advice from the system even when I don't know for certain that it is correct.
*F2 - When I am uncertain about a decision I believe the system rather than myself.
F3 - If I am not sure about a decision, I have faith that the system will provide the best solution.
F4 - When the system gives unusual advice I am confident that the advice is correct.
F5 - Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.

6. Personal Attachment

P1 - I would feel a sense of loss if the system was unavailable and I could no longer use it.
P2 - I feel a sense of attachment to using the system.
P3 - I find the system suitable to my style of decision making.
P4 - I like using the system for decision making.
P5 - I have a personal preference for making decisions with the system.

*Items used for adapted trust scale

ADAPTED TRUST SCALE

The tank spotting aid performs reliably.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

I understand how the tank spotting aid will assist me with decisions I have to make.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

When I am uncertain about a decision I believe the tank spotting aid rather than myself.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

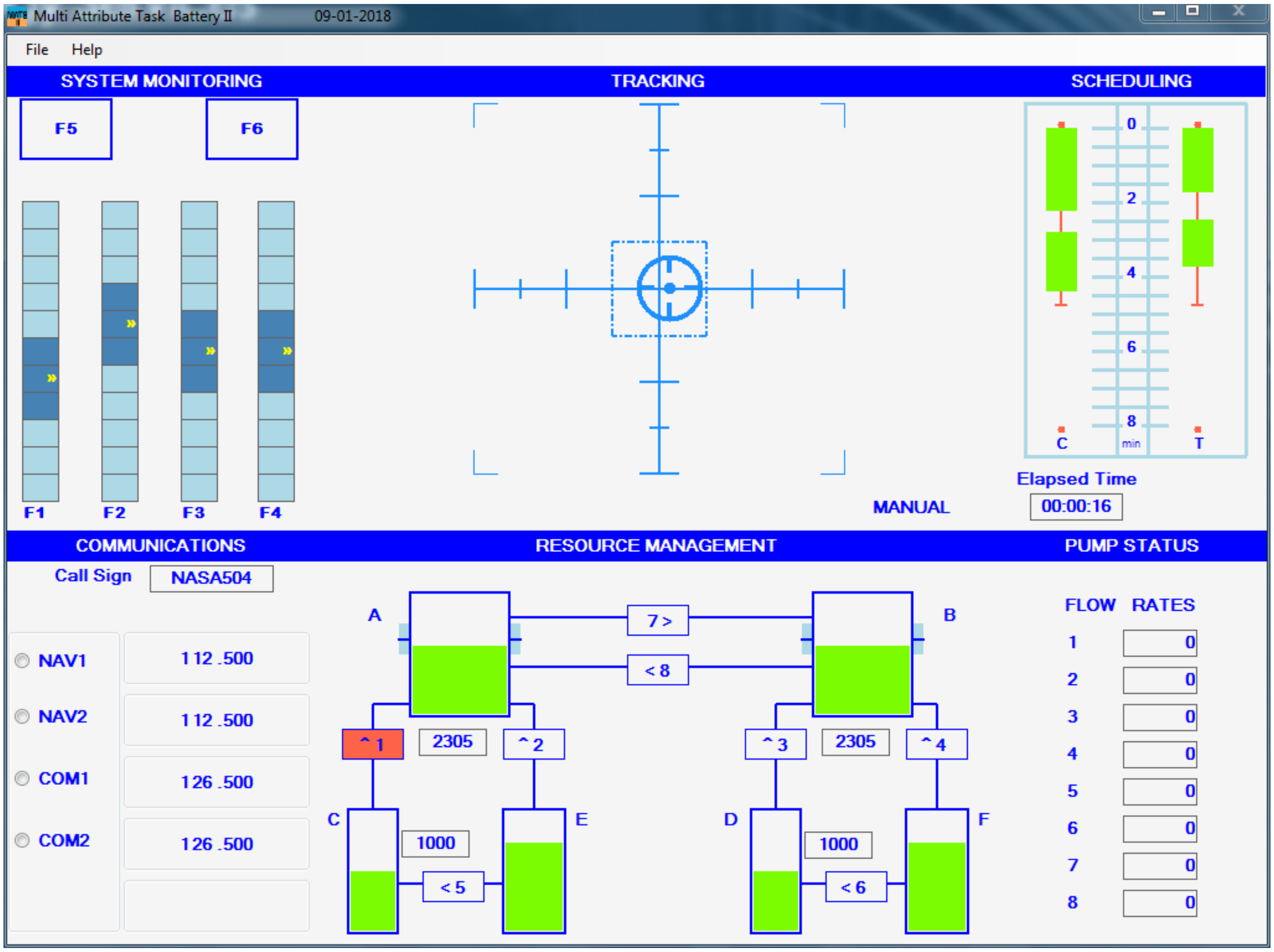
I can rely on the tank spotting aid to function properly.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

Although I may not know exactly how the tank spotting aid works, I know how to use it to make decisions about the problem.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

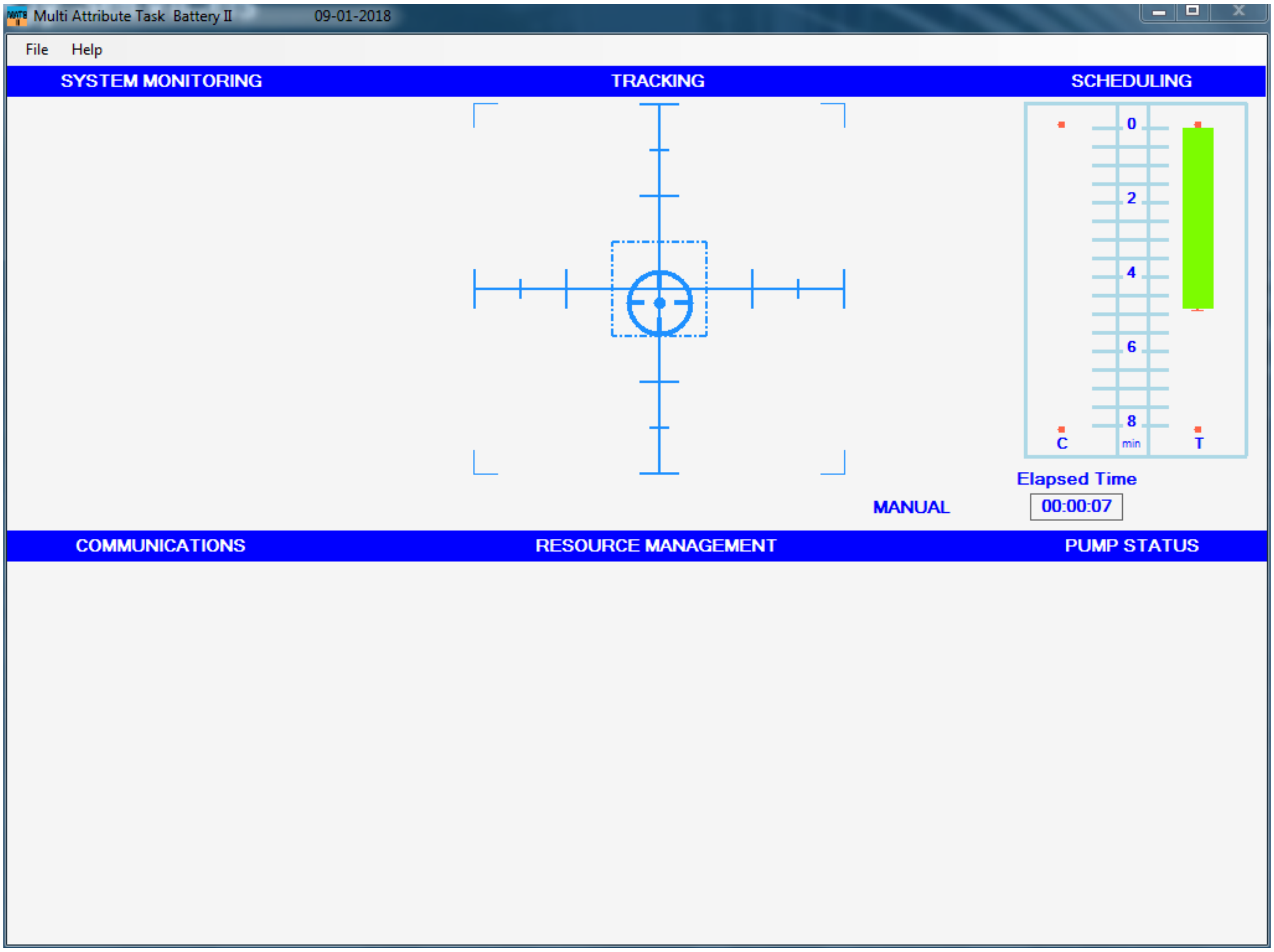
I believe advice from the tank spotting aid even when I don’t know for certain that it is correct.
Not descriptive: 1 2 3 4 5 6 7 8 9 10 11 12 : Very Descriptive

APPENDIX G

MATB II IMAGES



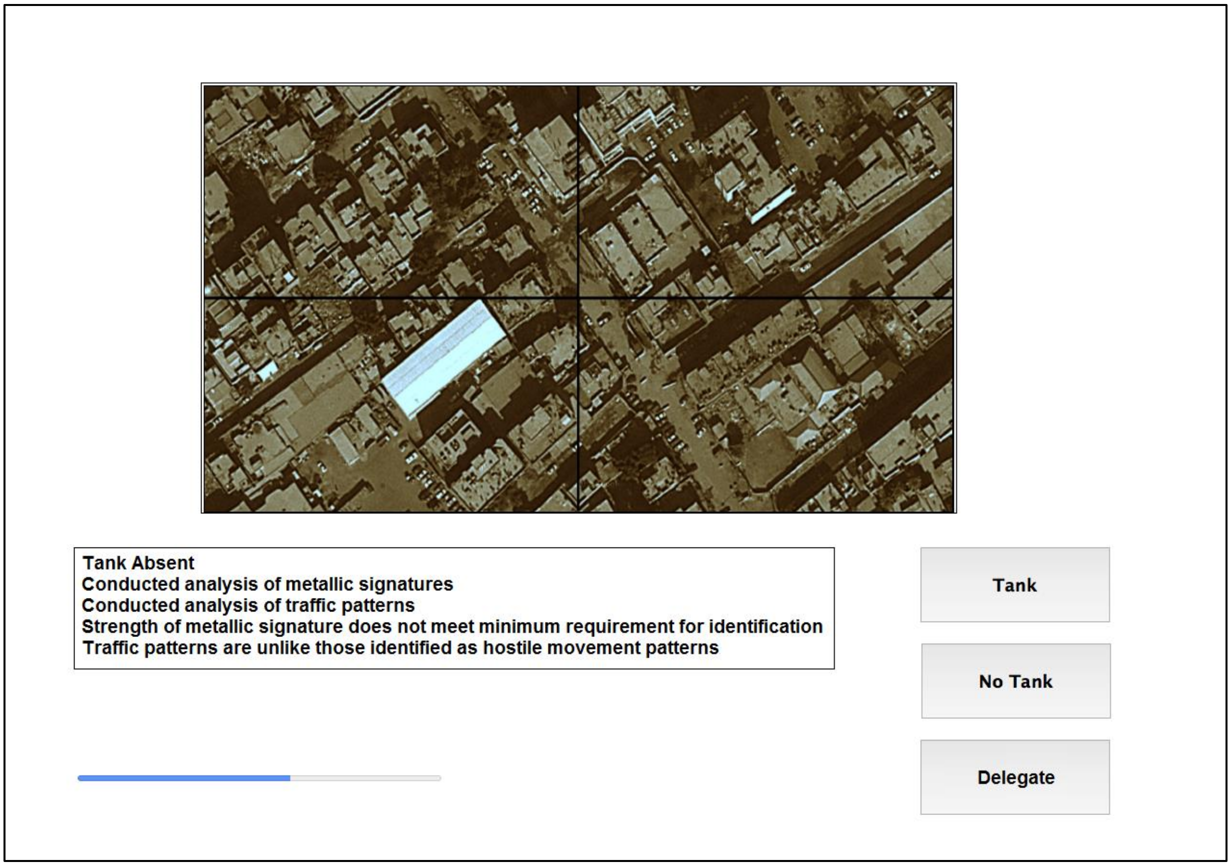
Screenshot of MATB II with all tasks



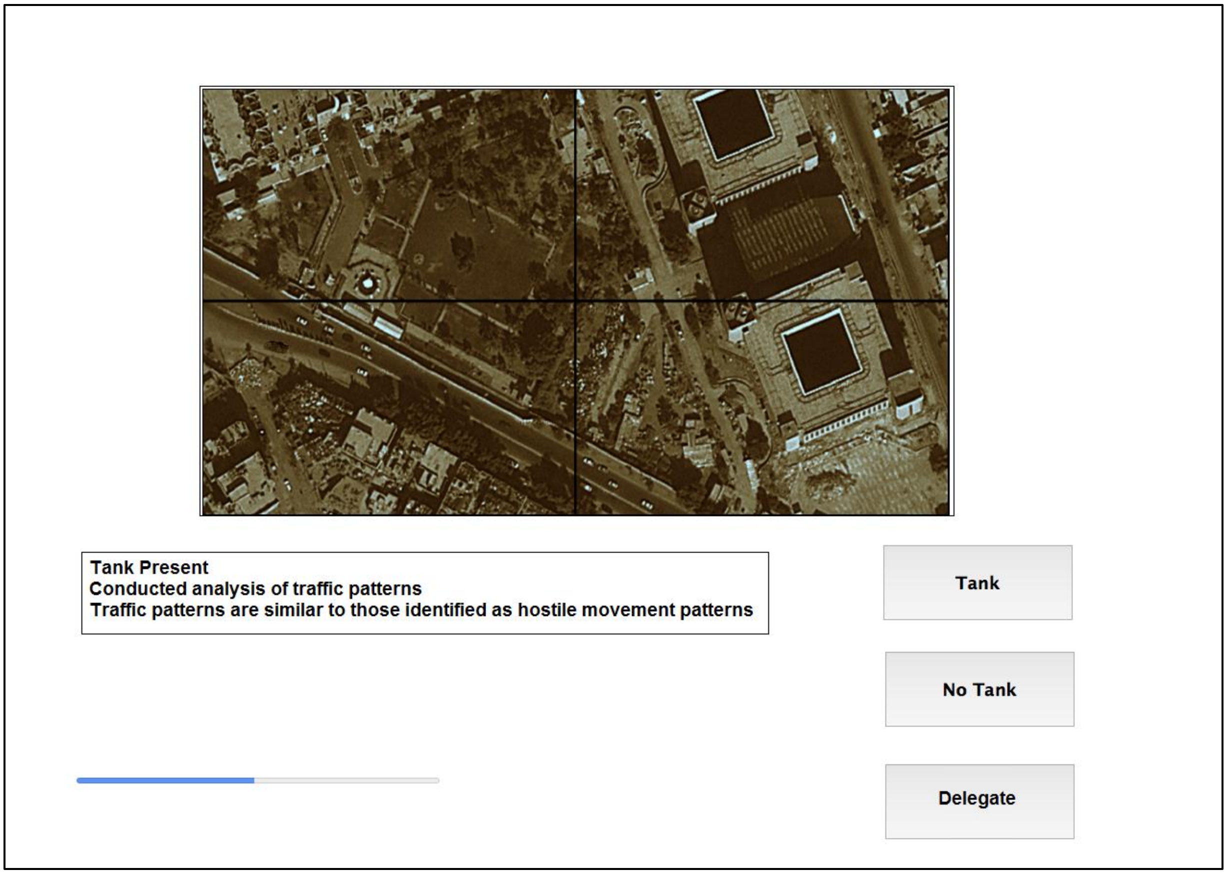
Screenshot of MATB II in the low workload setting, showing only the compensatory tracking task which will be used in the proposed study.

APPENDIX H

EXAMPLE TANK SPOTTING IMAGES



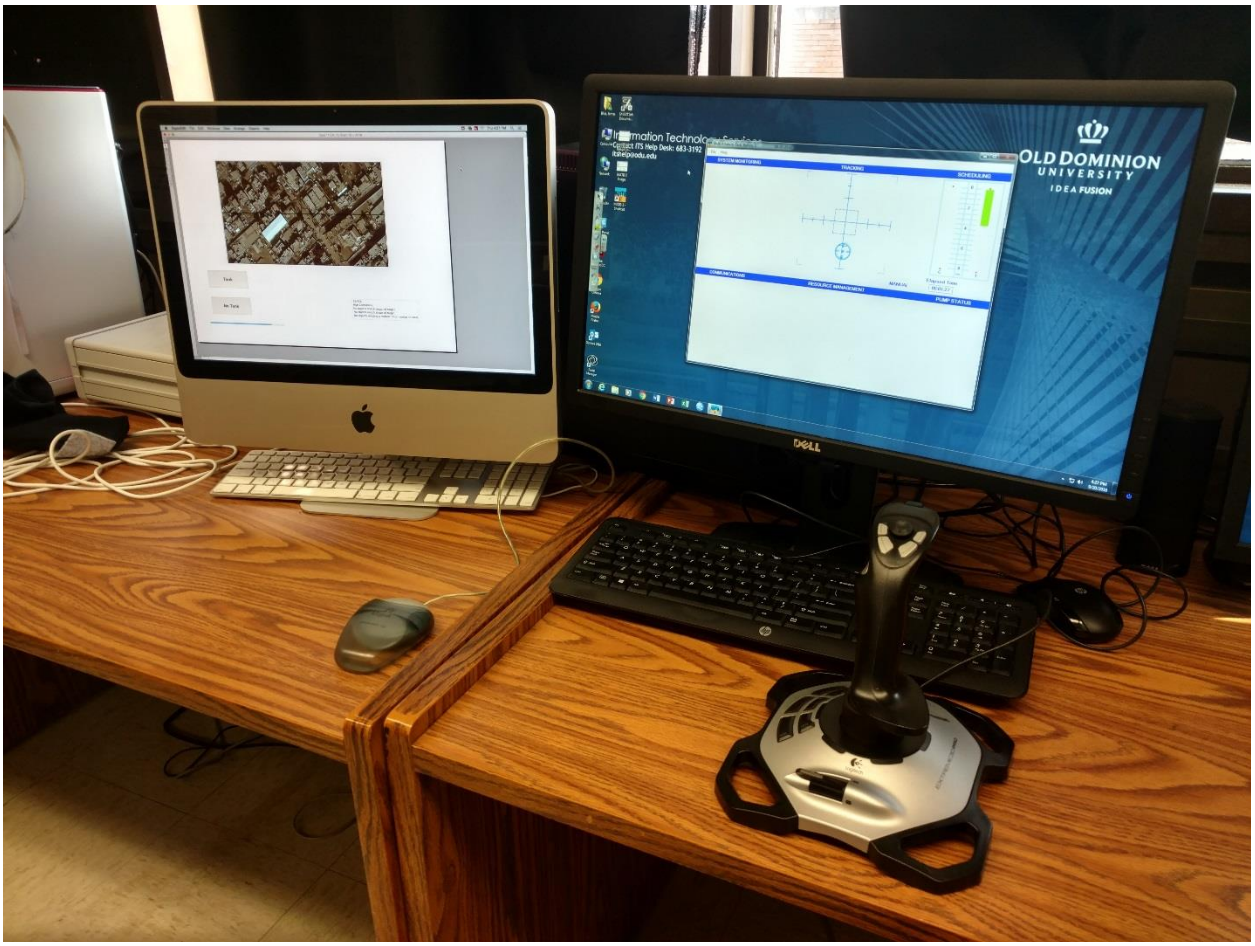
Example Tank Absent trial in high transparency condition. Screen contains transparency information, time tracking bar, tank and no tank buttons, and the delegate button to task shed.



Example Tank Present trial in low transparency condition. Screen contains transparency information, time tracking bar, tank and no tank buttons, and the delegate button to task shed.

APPENDIX I

EXPERIMENTAL SETUP



APPENDIX J

INFORMED CONSENT

INFORMED CONSENT DOCUMENT
OLD DOMINION UNIVERSITY

PROJECT TITLE: The Effects of Automation Transparency and Reliability on Task Shedding and Operator Trust

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 study, The Effects of Automation Transparency and Reliability on Task Shedding and Operator Trust will be conducted in Mills Godwin Building room 328.

RESEARCHERS

James P. Bliss, Ph.D., Full Professor, College of Sciences, Psychology Department, Responsible Project Investigator
William Lehman, Graduate Student, College of Sciences, Psychology Department

DESCRIPTION OF RESEARCH STUDY

Several studies have been conducted looking into the subject of automation trust and how operators use automation. None of them have explained how automation reliability and information explaining what the automation is doing can jointly influence users' trust as well as use of automation.

If you decide to participate, then you will join a study involving research using a flight tracking simulator as well as searching for a target within a map. You will use a joystick to control the flight tracking task. You will also view an image of a map and will use a mouse to choose whether there is a target in the map or not, or to give this task to an automated decision aid. You will be asked to fill out some brief questionnaires as well. If you say YES, then your participation will last for 60 minutes at the Mills Godwin Building room 328 at Old Dominion University. Approximately 80 other participants will be participating in this study.

EXCLUSIONARY CRITERIA

To participate in this study, you must be age 18 or over and must not have active duty military experience. To the best of your knowledge, you should not have participated in the Sona study ON-Tank Spotting that would keep you from participating in this study.

RISKS AND BENEFITS

RISKS: If you decide to participate in this study, then you may face minimal eye strain from normal computer use. The researcher tried to reduce these risks by restricting the study length to no more than 60 minutes. And, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: There are no known benefits from this study.

COSTS AND PAYMENTS

The researchers want your decision about participating in this study to be absolutely voluntary. Yet they recognize that your participation may pose some time inconvenience, therefore you will receive 1.5 ON-campus Sona credits which may be applied toward course requirements or extra credit for some Psychology courses. Equivalent credits may be obtained in other ways. You do not have to participate in this study, or any Psychology Department study, in order to obtain this credit.

NEW INFORMATION

If the researchers find new information during this study that would reasonably change your decision about participating, then they will give it to you.

CONFIDENTIALITY

The researchers will take reasonable steps to keep private information, such as questionnaires and performance data confidential. The researcher will remove any identifiers from the information. All data will be stored in a locked storage cabinet in the Psychology Department. The results of this study may be used in reports, presentations, and publications; but the researcher will not identify you. Of course, your records may be subpoenaed by court order or inspected by government bodies with oversight authority.

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. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation.

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 harm 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 investigators at the following phone numbers, Dr. James P. Bliss 757-683-4051, Dr. Tancy Vandecar-Burdin the current IRB chair at 757-683-3802 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 signing 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:
Dr. James P. Bliss 757-683-4051
William E. Lehman 906-284-2722

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. Tancy 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 signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

Subject's Printed Name & Signature	**Date**

INVESTIGATOR'S STATEMENT

I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

Investigator's Printed Name & Signature	**Date**

VITA

William Everett Lehman

Psychology Dept. Old Dominion University, Norfolk, VA 23529
welehman@gmail.com | linkedin.com/in/william-lehman-18917664/

Education

Old Dominion University — Norfolk, VA
M.S. Psychology — Expected August 2019
Human Factors Psychology Certificate in Modeling and Simulation — May 2018

Michigan Technological University — Houghton, MI
B.S. in Psychology, Law & Society minor, *magna cum laude* — May 2014

Recent Research Experience

Graduate Research Assistant, Old Dominion University
Research Environment for Alarm and Complex Task Simulation Lab — August 2016 – August 2019
Advisor: James Bliss, Ph.D.

Old Dominion University Research Foundation — May 2017 – December 2017
Supervisor: James Bliss, Ph.D.
Funding Agencies: U.S. Navy, Office of Secretary of Defense

Humane Interface Design Enterprise — May 2012 – May 2014
Michigan Technological University
Company: Chrysler

Undergraduate Research Assistant, Michigan Tech — June 2013 – June 2014
Mind, Music, Machine Lab
Advisor: Myeonghoon "Philart" Jeon, Ph.D.

Recent Professional Experience

Norfolk Botanical Garden — May 2018 – July 2019
Visitor Services Assistant

Sona Research Participation Advisor — August 2016 – August 2017

Michigan Tech Central Ticketing Office — January 2015 – July 2016
Office Assistant

Teaching Experience

Graduate Teaching Assistant, Old Dominion
Professor: Suzanne Morrow, M.S. — January 2018 – May 2019
Classes: Adolescent Psychology, Health Psychology, Developmental Psychology