Spring 2013

The Role of Trust as a Mediator Between System Characteristics and Response Behaviors

Eric T. Chancey
Old Dominion University

THE ROLE OF TRUST AS A MEDIATOR BETWEEN SYSTEM

CHARACTERISTICS AND RESPONSE BEHAVIORS

by

Eric T. Chancey
B.S. May 2010, Old Dominion University

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

MASTER OF SCIENCE

HUMAN FACTORS PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2013

Approved by:

_____
Dr. James P. Bliss (Director)

_____
Dr. Poornima Madhavan (Member)

_____
Dr. Robin J. Lewis (Member)

ABSTRACT

THE ROLE OF TRUST AS A MEDIATOR BETWEEN SYSTEM CHARACTERISTICS AND RESPONSE BEHAVIORS

Eric T. Chancey
Old Dominion University, 2013
Director: James P. Bliss

There have been several theoretical frameworks that acknowledge trust as a prime mediator between system characteristics and automation reliance. Some researchers have operationally defined trust as the behavior exhibited. Other researchers have suggested that although trust may guide operator response behaviors, trust does not completely determine the behavior and advocate the use of subjective measures of trust. Recently, several studies accounting for temporal precedence failed to confirm that trust mediated the relationship between system characteristics and response behavior. The purpose of the current work was to clarify the roles that trust plays in response behavior when interacting with a signaling system. Forty-four participants interacted with a primary flight simulation task and a secondary signaling system task. The signaling system varied in reliability (90% and 60%) within subjects and error bias (false alarm prone and miss prone) between subjects. Analyses indicated that trust partially mediated the relationship between reliability and agreement rate. Trust did not, however, mediate the relationship between reliability and reaction time. Trust also did not mediate the relationships between error bias and reaction time or agreement rate. Analyses of variance generally supported specific behavioral and trust hypotheses, indicating that the paradigm employed produced similar effects on response behaviors and subjective estimates of trust observed in other studies. The results of this study indicate that other mediating

variables may offer more predictive power in determining response behaviors. Additionally, strong assumptions of trust acting as the prime mediator and operationally defining trust as a type of behavior should be viewed with caution.

This thesis is dedicated to my incredibly supportive wife and parents.

ACKNOWLEDGMENTS

I would like to acknowledge my family, friends, and ODU faculty that have contributed to my academic journey. I would like to thank Jim Bliss for his continued guidance and for providing me with insight, challenges, and support when needed. Special thanks to committee members Poornima Madhavan and Robin Lewis for their valuable comments and advice that has greatly improved the quality of the ideas presented here. Thank you to past and present lab members for your friendship and support. Finally, to my parents who have always supported my endeavors and my wife who continues to be my greatest motivator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

One of the most prevalent factors thought to guide how operators depend on an automated system is the degree to which they trust that system. Some researchers, however, have approached the behavior of operator reliance and trust as being more or less the same, at least for measurement purposes (e.g. Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Meyer, 2001; Rice, 2009). For example, Meyer (2001) suggested that assessing operator behavioral response criteria could be a better alternative to measuring trust in simple warning systems than assessing verbal responses. In applied settings, administering questionnaires or interviews during a task is often not feasible, as it can be distracting and sometimes dangerous. Therefore, inferring trust from behavior could offer a powerful tool in regards to a reduction of research invasiveness. Some researchers, however, have suggested that inferring such relationships may not always be appropriate (Bustamante, 2009; Lee & See, 2004; Wiczorek & Manzey, 2010; Wiegmann, Rich, & Zhang, 2001).

Allowing trust to be operationally defined as a type of behavior confounds the final behavioral measure of operator trust with other constructs, such as self-confidence and workload, that would also undoubtedly affect behavior (Lee & See, 2004). To illustrate this point with an anecdotal example, a driver is traveling to an unfamiliar location navigating via global positioning system (GPS). The mapping function of the GPS has not been updated and frequently directs the driver to take inefficient routes. Perhaps the device indicates that the destination point is on the left side of a 6-lane road when in fact it is on the right side and 50 yards farther down the street. Although this system is unreliable, the user may still rely entirely on it, because of his or her

unfamiliarity with the environment. However, because of the GPS's inaccuracies, the user may still develop distrust for the system (see Ma & Kaber, 2007, which used route inefficiency to characterize GPS reliability). Therefore, although reliance upon an automated aid is closely related to the degree to which an operator trusts the aid, in some cases they may become dissociated. Trust is more appropriately designated as an attitude, a psychological component that contributes to the modification of a behavior (Lee & See, 2004). If a lack of trust exists between the operator and the automation, the modification of this behavior may be manifested by an absence of responses toward the automation or a degraded response rate (Breznitz, 1984). The purpose of the current work is to elaborate on the notion of trust in automation and empirically assess how the subjective measurement of trust mediates the behavior exhibited.

To provide guidance for what recommendations may be gathered from this research, it is important to ground the notion of an automated task in a theoretical context. Providing blanket recommendations for general automation, without the context of a theoretical framework, could be misguided and have the potential to generate performance decrements based on the type of automation implemented. For example, Sarter and Schroeder (2001) found a greater cost to pilot performance when an automated icing-monitoring system gave an incorrect course of action recommendation than a system that provided incorrect status information and left the course of action to the pilot. The authors noted that the performance differences might have reflected that each type of automation was aiding a separate stage of the decision process. Therefore, an objective of this research will be to provide automation recommendations that are specified by target system.

Automation

Automation is implemented for various reasons: the task may be too dangerous or even impossible for the human to accomplish; the task may be too difficult or unpleasant for the human; the automation may extend the human's capability; or a task is automated simply because it is possible (Wickens, Lee, Liu, & Becker, 2004, pp. 419-420). Regardless of the reason for automation, modifying a task with technology engenders certain consequences. Parasuraman, Sheridan, and Wickens (2000) propose that automating a task does not remove the human from the task, but rather modulates the human's involvement. Undeniably, it is now common to see tasks that were once performed entirely by the human being partially or fully undertaken by an automated component. The human is still, however, a vital element in these systems because the designer cannot foresee all potential outcomes that may be encountered. This unpredictability requires the creativity and adaptability the human possesses, to guide and monitor the system (Parasuraman & Riley, 1997). With the inclusion of the human component, however, performance cannot be predicted by the functionality of the tehcnology alone. Automation is, therefore, often evaluated by its influence on human behavior and performance (Parasuraman & Riley, 1997; Parasuraman et al., 2000; Parasuraman & Wickens, 2008).

In addition to describing *why* automation is implemented, it is also important to describe what it does. One perspective is to refer to automation as something that replaces or augments an aspect of the human information processing system (Parasuraman et al., 2000; Wickens et al., 2004, p. 420; Wickens & Hollands, 2000, p.

540). To wit, Parasuraman et al. (2000) proposes a four-stage model of automation based

on a simplified version of human information processing (Figure 1).

| Sensory Processing | Perception / Working Memory | Decision Making | Response Selection |
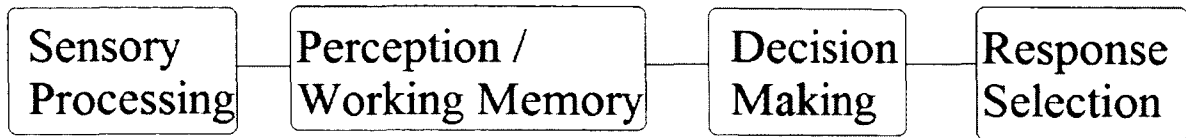
Figure 1. Simplified model of human information processing system. Adapted from *"A Model for Types and Levels of Human Interaction with Automation"* by R. Parasuraman, T. B. Sheridan, and C. D. Wickens, IEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30 (3), p. 287. *Copyright 2000 by IEEE.*

The four stages of automation are broken into: 1) information acquisition; 2)

information analysis; 3) decision and action selection; and 4) action implementation. The

authors also provide a definition for automation as being "a device or system that

accomplishes (partially or fully) a function that was previously, or conceivably could be,

carried out (partially or fully) by a human operator" (Parasuraman et al., 2000, p. 287).

This definition implies a gradient of human involvement with a task that may be

controlled by an automated component to some specified degree. The degree to which the

human is involved is referred to as the level of automation (LOA), and is characterized by

the trade-off between human operator and automation for overriding authority over a

course of action. A lower LOA indicates the computer has little authority over the course

of action and a higher LOA indicates that the computer has more authority and may even

act autonomously (for a similar framework see Endsley & Kaber, 1999).

As previously mentioned, it is important to specify the automation under

investigation to provide appropriate recommendations from which research may then be

generalized. The reason for drawing attention to this detail is to simply provide a theoretical framework from which to proceed. As implied from the previous discussion on "what automation does," there is a multitude of ways to implement automation to fully or partially (via the set LOA) replace stages of the human information processing system, which would be reflected by the stage of automation. So although theories of "general automation" may be pertinent across a wide selection of automation levels, it is also of interest to specify the level of automation under investigation to better specify when theoretical predictions, such as those motivated by trust in automation, would not generalize from one level to another. For example, it is reasonable to expect that trust would develop differently between low LOA systems and high LOA systems. With high LOA systems the technology would be almost entirely responsible for carrying out a task. The need for a high LOA may be due to the high complexity of the task or the human's inability to perform the task or perform it as well as the computer. The human, therefore, may have no other choice but to rely on the system. This could lead to a situation where the human totally trusts the automation or helplessly relies on the automation independent of his or her trust. Conversely, a system characterized by a lower LOA would to some degree timeshare the task with the human. The human, in this case, would be in a position to take over for the system if he or she disagrees with it. In cases of disagreement with the automation, trust may play more of a role in whether or not the human relies on the automation. The specific LOA under investigation in the current work is that of sensor-based signaling systems. The following section provides a more detailed discussion concerning this specific automation.

Sensor-Based Signaling Systems

Automation has become prevalent in applied settings such as aviation (Sarter & Woods, 1992, 2000), medicine (Krupinski, Nodine, & Kundel, 1993; Elkin et al., 2010), military (Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001; Wang, Jamieson, & Hollands, 2009), and industry (Drury & Sinclair, 1983). Because of the growing use and complexity of automation, the role of the human has progressively become that of a system monitor. To facilitate interactions with the potentially complex or numerous amounts of systems, the human is frequently presented with various sensor-based signaling systems that aid him or her in monitoring those tasks carried out by a computer (Bliss & Gilson, 1998).

The term, "sensor-based signaling systems" (which will subsequently be referred to as "signaling systems") was chosen to "represent a broad category of stimuli including warnings, alarms, and alerts" (Bliss & Gilson, 1998, p. 58). Also referred to as "dynamic warnings," these systems are designed to direct attention to potential hazards that may require closer inspection or intervention on the part of the user (Meyer, 2004). Returning to the cognitive framework proposed by Parasuraman et al. (2000), this type of automation, therefore, replaces the initial stage of information processing (see Figure 1) and may be classified as an example of Stage 1 automation (Wickens et al., 2004). Signaling systems have been implemented and studied in a multitude of domains including security monitoring (Bliss & Chancey, 2010), aviation (Pritchett, 2001), medical theaters (Meredith & Edworthy, 1995), ground transportation (Lees & Lee, 2007), power plants (Carvalho, dos Santos, Gomes, Borges, & Guerlain, 2008), and dismounted military operations (Dzindolet et al., 2001).

The simplest paradigm employed to investigate signaling systems includes two components: the mechanical device itself and the human operator (Getty, Swets, Pickett, & Gonthier, 1995; Bliss & Gilson, 1998). The first component, the signaling system, operates on conditional logic governed by a sensor threshold. If a condition exceeds the preset threshold then a signal is issued, for example a smoke alarm will issue an alert if enough smoke is present in the environment to exceed its preset threshold. If the sensor is set too conservatively, then false alarms will be minimized at the expense of dangers not being identified. If the system is set too liberally, then the potential for averting a dangerous condition will be minimized at the expense of an increase in false alarms. Designers of these systems should attempt to strike a balance between these two extremes, as they affect the attitudes and behaviors of the second component, the human (Bliss & Gilson, 1998).

To illustrate this point, if the system is set too liberally, producing frequent false alarms, this could garner a reduced or slowed response rate (Getty et al., 1995). More seriously, the user may ignore or disable the system altogether, even when it does signal a true event (Sorkin, 1988). This type of behavior has been termed the "cry-wolf" effect or phenomenon (Breznitz, 1984). Conversely, if the system is set too conservatively, the system may fail to signal potentially crucial events. This may reduce the user's reliance upon the automation, forcing the user to monitor the raw data. Generally, in applied settings, more often than not users must engage in multiple tasks at the same time. In these types of multi-task scenarios, operators will be required to divide attention which can lead to an increase of workload and a reduction of performance (Dixon & Wickens, 2006; Dixon et al., 2007). Because of the potential for false alarms and misses to be

8

introduced into this paradigm, and the subsequent effects these have on human behavior, system reliability is often discussed in the automation literature (see Wickens & Dixon, 2007).

Reliability and Automation Use

Signaling systems are regularly implemented under the erroneous assumption that the user will acknowledge the authenticity of the signal and respond appropriately and consistently (Bliss & Gilson, 1998). To the contrary, as mentioned in the previous section, research has shown that users exhibit inefficient response behaviors when presented with systems that are unreliable (Breznitz, 1984; Bliss & Gilson, 1998; Meyer, 2001) or perceived to be unreliable (Bliss, Dunn, & Fuller, 1995). The unreliability of the system may, therefore, lead to instances of user under-reliance. Conversely, however, if the system is perceived as being highly reliable or produces errors infrequently, the user may become complacent and not intervene when necessary. The result of the operator either over-relying or under-relying on automation has been characterized as automation *misuse* and *disuse* (Parasuraman & Riley, 1997). These are considered maladaptive automation reliance behaviors that have the potential to degrade safety and profitability (Lee & See, 2004).

Issues concerning the under and overreliance on systems have been investigated in various domains (Parasuraman & Riley, 1997). For example, a combat identification (CID) aid is a type of signaling system to help dismounted Soldiers designate other individuals collocated on the battlefield, via interrogator and transponder signals, as either "friendly" or "unknown." A more specific purpose of these aids is to reduce the occurrences of fratricide – the inappropriate allocation of lethal force to friendly

personnel (Webb & Hewett, 2010). Dizindolet et al. (2001) noted that if Soldiers were to perceive these devices as being unreliable, widespread instances of disuse could negate their utility. This would lead to wasted resources and an unaffected impact on fratricide rates. Conversely, if these systems were perceived as being highly reliable, instances of misuse could occur. In ideal conditions, these systems operate at above a 99% confidence level. In theater, however, these systems would likely operate in a state of fluctuating reliability as the battlefield is of a dynamic nature and other friendly Soldiers may not always have functioning transponders (Dzindolet et al. 2001). This could, therefore, lead to instances of fratricide via CID aid designation of friendly Soldiers as being "unknown." However, it should be noted that to the knowledge of this author, at the time of writing this document, these systems have not been evaluated in theater and, therefore, representative reliability ranges have not been documented.

The ubiquity of automation over- and under-reliance provides researchers a framework from which to identify common factors present in either overuse or underuse of automation. Identifying these factors has implications for training automation users and designers to be aware of instances in which the human-automation partnership may break down. Furthermore, policies and procedures may be adapted (Parasuraman & Riley, 1997), limited resources may be allocated more appropriately (Dizindolet et al., 2001), and, as a result, more informed decisions may be reached as to whether or not automation should be implemented. One of the most prevalent factors thought to guide user reliance is user trust (Sheridan & Hennessy, 1984; Muir, 1987; Parasuraman & Riley, 1997; Meyer, 2001; Lee & See, 2004; Madhavan & Wiegmann, 2007; Rice, 2009).

The subsequent section provides an overview of trust in automation and relates this construct to the idea of user reliance and response behaviors.

Trust in Automation

The notion of trust in automation is an important construct in the evaluation of automation reliance. To illustrate this point, several theoretical frameworks have been proposed in which trust is a central component in this relationship (Muir, 1987; Lee & See, 2004; Parasuraman & Riley, 1997; Dzindolet et al., 2001; Wiegmann et al., 2001; Madhavan & Wiegmann, 2007). Lee and See (2004) have suggested that the impetus for investigating trust in this context is rooted in the movement of technology becoming increasingly cognitively complex. Similar to the role trust plays in human-human interactions, the authors suggest that trust in automation reduces the need for continuous supervision and increases decision certainty. Indeed, generalizing from the social psychological perspective of trust among humans would provide a logical foundation from which theories of human trust in automation could be proposed.

To illustrate this generalization, Nass and colleagues have proposed the "Computers Are Social Actors" (CASA) paradigm, which postulates that the social rules that guide human interactions also guide the interactions between humans and computers (Nass, Steuer, & Tauber, 1994; Nass, Moon, Fogg, Reeves, & Dryer, 1995; Nass & Moon, 2000; Nass & Lee, 2001). Within this line of research the authors have shown that participants demonstrate gender stereotypes toward computers (Nass, Moon, & Green, 1997), politeness toward computers (Nass, Moon, & Carney, 1999), reciprocal self-disclosure toward computers (Moon, 2000), and "similarity-attraction," where

participants favored computers matching their own personalities (Nass et al., 1995; Nass & Lee 2001; for review of this research see Nass & Moon, 2000).

Lee and See (2004) suggest that, based on some of the CASA research and other such similar work (Miller, 2002; Norman, Ortony, & Russell, 2003), the emotional and attitudinal aspects (including trust) that mediate social relationships among humans may also mediate the relationship between humans and technology (i.e. automation). Indeed, this is not an entirely novel concept, as Halpin, Johnson, and Thornberry (1973) referred to instances of "mistrust" and "faith" in this respect as well. Furthermore, Sheridan and Hennessy (1984) have suggested that the user's perception of a system's "trustworthiness" will facilitate predictions of performance.

There have been several frameworks that acknowledge trust as a prime mediator between system characteristics and automation reliance. Many of these frameworks share overlapping qualities and are, in some cases, updated versions of previous models. The following subsections will provide an overview of several conceptualizations of trust in automation. Depending on which perspective a researcher subsumes, some measurement techniques will be considered more appropriate indices of trust than others (i.e. behavioral responses versus subjective responses).

Behavioral Inferences of Trust

Although several researchers have used subjective ratings to assess user trust in automation (e.g. Lee & Moray, 1992, 1994; Moray, Inagaki, & Itoh, 2000; Wang et al., 2009), Meyer (2001) proposed that recording differential setting of user response criterion (i.e. responder's β in a SDT paradigm) could serve as a viable alternative to assessing trust subjectively, when interacting with a signaling system. The author

suggests that the advantage of assessing operator trust by recording behavioral responses is that it allows the researcher to monitor trust development over time without the invasiveness that subjective questionnaires may require.

From this perspective, Meyer (2001) conducted a study in which he varied the validity of a warning display that indicated system state (system OK if indicator is green, failure if red). The results of this study indicated that participants adjusted their responses to the validity of the warnings. However, participants tended to respond to the red warnings as though there was a failure more often than if the indicator was green.

Based on these results, the author proposed the need for a "refinement" of trust. More specifically, he suggested that there are two different types of trust, and that they are affected by separate independent variables: 1) when a signaling system indicates that there is a critical condition present and issues an alert 2) when a signaling system does not deem a critical condition is present and does not issue an alert. He refers to the behavior associated with each type of trust as compliance and reliance.

Compliance refers to the response of the operator when a signaling system indicates that there is a danger and issues a warning (i.e. behavior exhibited when the indicator was red). The operator is compliant if he or she responds to this warning as if there is an existing danger. If, however, the operator is not compliant then he or she will disregard the signaling system. Although, it is unclear where the cutoff is that would designate a delayed response as being non-compliant or compliant but simply delayed. The notion of action intent may provide an interesting avenue from which to evaluate compliance behaviors, however, it is beyond the scope of the current work.

Reliance refers to the response (or non-response) of the operator to a signaling system when it indicates that the state of the system is within normal conditions (i.e. behavior exhibited when the indicator was green). If the operator assumes that the signaling system is correctly indicating a safe state, then he or she will continue without response or, in the case of a system that issues an OK signal, he or she will respond by agreeing with it. The operator may also, however, not rely on this system and manually or visually inspect to certify that the system is correctly indicating a safe state. It should be noted that, besides in contrived laboratory conditions, it is difficult to identify real world systems which indicates an OK state that requires a response by the operator. Generally, systems that indicate an OK state would motivate the operator to take no action. However, although far less common than warning systems that would signal a potential problem, there are signaling systems that would issue an OK signal. For example, the "straight board" of a submarine indicates the hull opening status of the vessel. Prior to submersion, gauges must be consulted to determine that all openings are closed – this would be an indication of an OK state to dive. If the crew is reliant on these gauges, then they would respond by diving, if they were not reliant on these gauges they would inspect hull openings in spite of the OK advisory.

Importantly, the term "reliance" in this context needs some clarification. Parasuraman and Riley (1995) use the term reliance to refer to the degree to which an individual engages or disengages automation. Their usage of the term reliance, however, does not necessarily represent the specific behavior adopted by a user when a signaling system indicates a normal state of operation. Some researchers have adopted the term *dependence* to refer to behavioral measures of automation usage rates, compliance, and

reliance (Dixon & Wickens, 2006; Dixon et al., 2007; Rice, 2009). To keep these two concepts of reliance separated, the current work will use the term dependence to denote compliance, reliance, and the degree to which an individual engages or disengages an automated component.

When a warning is issued, the actual environmental state may be signal present (hit) or signal absent (false alarm). Therefore, false alarms would be associated with compliance behaviors. When a warning is not issued or the system indicates an OK state, the actual environmental state may be signal absent (correct rejection) or signal present (miss). Therefore, misses would be associated with reliance behaviors. Rice (2009) suggests that "in its strongest theoretical form" (p. 306), Meyer's (2001, 2004) conceptualization of the process that mediates the relationship between the signaling system and the subsequent responses, resembles the model in Figure 2B. In this depiction, false alarms affect compliance and misses affect reliance via independent processes. It should be noted, however, that Meyer's 2004 article makes only one mention of trust as playing any role in operator compliance or reliance. Meyer instead lists trust in automation as one of multiple inputs that are subsumed under "system specific characteristics." More specifically, he focuses on three broad categories of influences on operator responses: normative, task, and operator factors. It is unclear if Meyer (2004) abandons the notion of two separate types of trust, or if he is simply clarifying the underlying characteristics (including trust in automation) that affect reliance and compliance differently. In either case, researchers have proceeded from this initial perspective and assumed that trust is still the main underlying factor affecting both outcome behaviors (e.g. Dixon & Wickens, 2006; Dixon et al., 2007; Rice, 2009).

A

False-alarm rate		Compliance
Trust	
Miss rate		Reliance

B

| False-alarm rate | → Trust in alerts → | Compliance |
| Miss rate | → Trust in nonalerts → | Reliance |

C

| False-alarm rate | → Trust in alarms → | Compliance |
| Miss rate | → Trust in nonalerts → | Reliance |

D

| False-alarm rate | → Trust in alarms → | Compliance |
| Miss rate | → Trust in nonalerts → | Reliance |
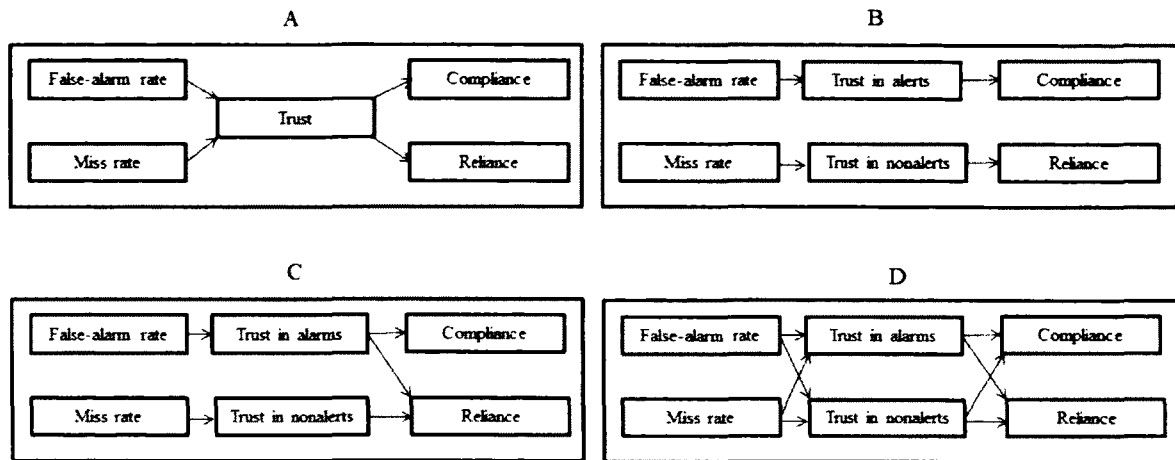
Figure 2. Signaling system and the subsequent behavior mediated by trust in several process models. Adapted from "Examining single- and multiple-process theories of trust in automation," by S. Rice, *The Journal of General Psychology*, 13(3), p. 307. Copyright 2009 by Heldref Publications.

Building on the perspective of compliance and reliance, several studies have investigated the independence of the two processes underlying both behaviors. Using a simulated unmanned aerial vehicle (UAV) task, Dixon and Wickens (2006) conducted two experiments in which pilots flew a mission while also monitoring a set of gauges that would occasionally indicate a failure. To help the pilot monitor gauges, a diagnostic aid alerted the participants to a system failure. To test the independence of compliance and reliance, in both experiments two of the aids had bias error rates so that one was more likely to issue a false alarm and the other was more likely to miss a failure. Collapsing across experiments, the authors found that miss rate did not correlate with system failure reaction time or detection rate in response to the diagnostic aid, which would be indices of compliance. Also, false alarm rate did not significantly correlate with indices of

reliance such as reaction time to system failure misses by the diagnostic aid and secondary task performance measures, which would cause participants to divert attention to the gauges to monitor the raw data. This evidence would support the model in figure 2B. However, the authors note that the correlations of false alarm rate and reliance indices were approaching significance.

Building off of this study, Dixon et al. (2007) used a dual task paradigm where participants were again aided by false alarm prone (FP) or miss prone (MP) diagnostic automation. As opposed to their previous endeavor, this study did find that FP automation affected both compliance and reliance, which provided support for the non-independence of compliance and reliance. The MP automation, however, seemed to only affect reliance behaviors, suggesting figure 2C as the most adequate model.

Finally, using a UAV simulation paradigm where participants were aided by a diagnostic aid in the identification of enemy tanks over Baghdad, Iraq, Rice (2009) provided evidence for a multi-process model (figure 2D). The diagnostic aid was, again, either FP or MP and varied in reliability from 95% to 55% in 5% increments. To differentiate between the model in figure 2A (single process) and 2D (multi-process), the author conducted a state-trace analysis that revealed a nonmonotonic relationship between the variables, whereby the type of bias (FP and MP) did not create equal increases in agreement rates and response times from the FP and MP automation conditions.

Within these frameworks, researchers have inferred user trust from the behavior exhibited (Dixon & Wickens, 2006; Dixon et al., 2007; Meyer, 2001; Rice, 2009). Although Rice (2009) states that, "*Trust* refers to a psychological state, whereas

dependence refers to a behavioral measure," (p. 304, italics in original text) within this study he infers trust by assuming that "… when participants trusted the automation, they would quickly agree with the aid" (p. 312). He also notes, however, that dependence and trust are not always perfectly correlated. He provides the example of an overloaded operator who may need to depend on a system because of inadequate time to ensure its accuracy.

All of the aforementioned studies have provided valuable insight into the investigation of changes in operator dependence as a function of a signaling system's characteristics. Within this research the authors frequently reference trust as an explanation for the outcomes. It may be that these authors are relatively safe in using this construct as the general cognitive prerequisite for operator dependence. However, some researchers advocate caution in regard to inferring trust from behavior. The following section will elaborate on this perspective.

Trust as a Partial Mediator

Lee and See (2004) provide a review of trust in automation, in which they state that "trust guides – but does not completely determine – reliance…" (p. 51). The authors are referring to reliance in the sense of what the current work has termed dependence. Within the review, they attempt to clarify inconsistencies that have emerged from research concerned with trust in automation. Pertinent to the current focus on trust in automation are definitional inconsistencies that have implications for how trust may then be operationally defined.

Researchers have characterized trust from several different perspectives, as a belief, an attitude, an intention, or as a behavior. The authors attempt to reconcile these

inconsistencies by utilizing the framework developed by Ajzen and Fishbein (1980).

From this perspective, behaviors occur as a result of intentions, intentions result from

attitudes, and attitudes are based on beliefs. Lee and See (2004) develop this idea by

suggesting that beliefs and perceptions are the product of an individual's experience or

available information. This provides the informational base to develop attitudes. An

attitude, then, acts as an affective evaluation of a belief and motivates the adoption of a

particular intention, which results in a behavior (See Figure 3). Within this framework,

the authors have separated behavioral outcomes, such as automation dependence, from

attitudes, such as trust, that may affect those intentions and behaviors. This

conceptualization of trust does not allow its effects to be confused with other factors that

may influence behavior such as workload, situation awareness, and self-confidence of the

operator (Lee & See, 2004, p. 53). The authors ultimately define trust as "the attitude that

an agent will help achieve an individual's goals in a situation characterized by

uncertainty and vulnerability" (Lee & See, 2004, p. 54).

Experience Availability of Information → Beliefs Perceptions → Attitudes → Intentions → Behavior

Affective evaluation of beliefs →

Dependence

Display Automation
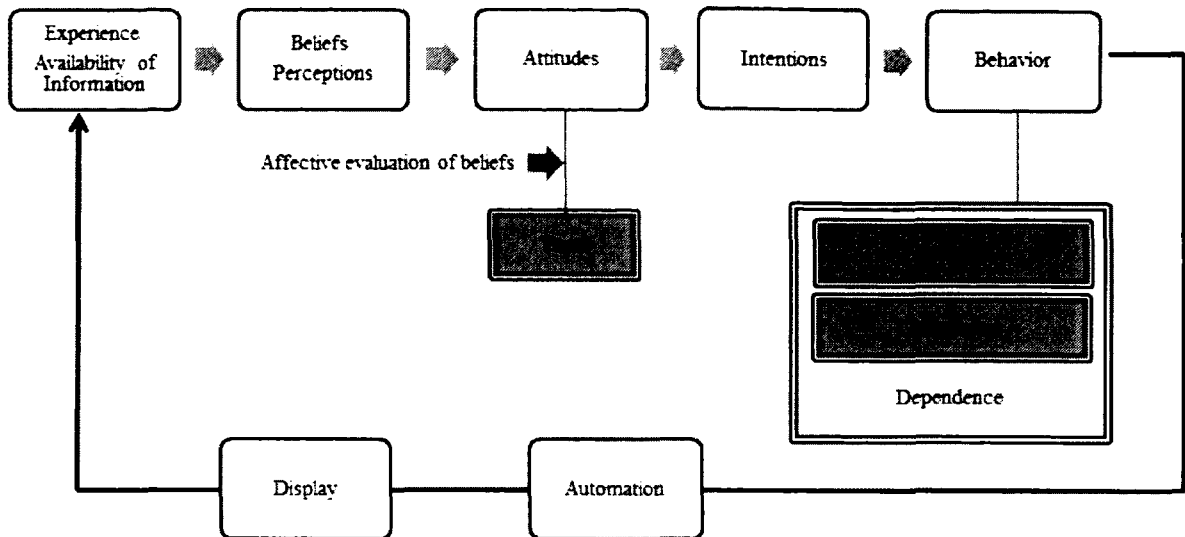
Figure 3. Simplified graphical representation of Lee and See's (2004) compartmentalization of beliefs, attitudes, intentions, and behavior with regards to trust and dependence

An important implication of this conceptualization of trust is that it does not

suggest trust is the only mediating variable between automation characteristics and

behavior. Indeed, the authors further note that many definitions of trust emphasize the

importance of the goal-oriented nature of the construct. With this in mind, trust is of

graded importance in situations where the trustee (automation) is less relevant in terms of

furthering a trustor's (operator) goals. This would imply that other (non-trust) factors

could offer more explanatory power in what motivates the subsequent behavior.

Using a simulated semi-automatic pasteurization plant, Lee and Moray (1992)

examined the role of trust in participant allocation of either automatic or manual control

in plant operation. Results of this study showed that trust failed to fully account for

changes in automation dependence under certain conditions. The authors suggested that

dependence strategies may have also occurred as a function of participant self-confidence in their own abilities. Subsequently, Lee and Moray (1994) used a similar paradigm which included subjective ratings of participant self-confidence. Results indicated that, in general, when self-confidence is greater than trust, then manual control is preferred. Conversely, if trust in the automation is greater than self-confidence, then dependence on the automation is preferred. Thus, trust alone would not seem to influence operator behaviors.

Along a similar perspective, Wiegmann et al. (2001) noted the importance of separating behavioral measures of dependence from subjective measures of trust. They suggested that automation trust should be defined in terms of subjective measures of confidence or estimates of reliability and that automation dependence should be defined in terms of performance and behavioral measures. The authors further suggest two instances in which an operator's use of automation may become dissociated from levels of trust. First, the automation may be unreliable but still be more accurate than the operator. Second, the operator may lack the information needed to inform a diagnosis.

From this, Wiegmann et al. (2001) conducted a study in which participants utilized an automated aid that was (unbeknownst to the participant) 60%, 80%, or 100% reliable at diagnosing a system failure. In the 60% reliability condition the aid would raise to 80% half way through the experimental session, whereas the 100% reliable aid would lower to 80% half way through the experimental session. The 80% reliable aid remained constant throughout. The authors inferred trust from the participants' subjective estimates of the aid's reliability and confidence in the aid. Objective measures of performance were inferred from agreement with the aid and decision time.

Correlations among the performance and subjective variables revealed mixed support for inferring trust from behavior. Subjective estimates of reliability consistently correlated with agreement rates at the $p < .05$ for the pre- and post-shift of aid reliability (see Table 1, p. 363). However, participant confidence in the aid was not significantly correlated with agreement rate per reliability shift except between the pre-shift measurements (confidence and agreement $r = .296$, p < .05). Reaction time for agreeing with the aid did not significantly correlate with any of the subjective measures pre- or post-shift. The authors suggest a degree of caution be taken to distinguish between subjective measurements of trust as a psychological construct and performance measures as an indication of dependence.

Two factors of these results are of interest in relation to the current work. First, Rice (2009) infers trust from participants quickly agreeing with a signaling system. Wiegmann et al. (2001), however, did not find that the subjective estimates of reliability and confidence (i.e. trust) correlate with reaction time. They did, however, find that reliability estimates consistently correlated with agreement rates and, at least in the pre-shift trial, also correlated with confidence ratings. This may or may not provide evidence that Rice's (2009) assumption of trust is at least partially tenuous (based on Wiegmann et al.'s particular data set).

Wiegmann et al. (2001) approached trust as a subjective estimate of aid reliability and confidence in that aid. This view of trust, although, may not fully encompass the multidimensionality of the construct. However, Wickens and Hollands (2000) have conceptualized trust from the perspective of perceived reliability in relation to actual reliability in reference to trust calibration. To address this perspective in light of another

framework, Lee and See (2004) have suggested three general bases for trust in automation: performance, process, and purpose. Within this framework, reliability would fall under performance, which refers to what automation does and to what degree it competently contributes to achieving the user's goals. From this, perceived reliability would seem to contribute to the measurement of trust, as it is based in performance.

To address the use of subjective confidence in the system, trust has been defined as a "willingness to rely on an exchange partner in whom one has confidence" (Moorman, Deshpande, & Zaltman, 1993, p. 82). This perspective, however, would seem to put the measure at odds with the definitional compartmentalization subsumed by Lee and See (2004), because trust would now become an intention and not an attitude. However, this definition is from a social human-human perspective not a human-automation one. Also, Wiegmann et al. (2001) do not explicitly suggest a definition of confidence and, therefore, it may be independent of an intention. For example, the operator may have confidence in the automation's ability to perform an action (which would denote an attitude), but not necessarily confidence in its willingness to perform a task. Indeed, Jian, Bisantz, and Drury (2000) have empirically developed a measure of trust in automation that includes an item gauging operator confidence in automation. However, it may be precipitous to discount behavioral measures of trust based on two subjective indices. Therefore, this author agrees with Wiegmann et al.'s suggestion of caution in terms of the measurement of trust via behavior, but advocates a more multi-dimensional measure of trust.

Clearly, it is important to document circumstances that lead to misuse or disuse as well as parse out the deleterious effects of misses and false alarms individually. From the

perspective of the proposed research, however, the performance and dependence

behaviors investigated are not a primary concern. What is less clear is whether trust is the

prime mediating factor between the system characteristics and operator behavior. None of

the studies covered within the "Behavioral Inferences of Trust" subsection assessed trust

subjectively and instead assumed its role based on the behavior observed. Several recent

works, however, have suggested that trust may not mediate the relationship between

system characteristics and dependence behavior (Bustamante, 2009; Wiczorek &

Manzey, 2010). The following section presents an overview of these studies.

Trust does not Mediate Dependence

To this point in the review, it has been assumed that trust mediates the

relationship between system characteristics (i.e. FP versus MP and automation reliability)

and subsequent user dependence on that system to some degree. Some researchers have

assumed trust as the sole mediator, while others have advised caution in this assumption

and suggested that trust be measured subjectively to at least allow for other

"confounding" constructs. However, recently there have been at least two studies in

which trust was not found to mediate the relationship between system and dependence.

Bustamante (2009) noted that to adequately test the assumption of trust as a

mediating factor between system and dependence, two criteria of experimental design

must be met: association and temporal precedence. First, to establish a causal effect of

trust on dependence there must be an association between the construct and the behavior.

Indeed, based on the aforementioned theoretical models and research there is ample

evidence to satisfy this criterion (see Lee & See, 2004; Madhavan & Wiegmann, 2007 for

reviews). Bustamante (2009) suggested, however, that the second criterion of temporal

precedence has been neglected in the existing literature. Researchers tend to administer the measure of trust following the interaction with a system. This, as the author notes, creates the difficult task of inferring the directionality of the causal relationship if dependence behaviors are measured prior to the trust variable. It should also be noted that to thoroughly address the subject of causality, two other criteria must also be met to assume a causal relationship: 1) The relationship must be nonspurious or hold when the influence of other possible variables is eliminated and 2) there needs to be a theoretically plausible mechanism by which trust can influence dependence (Cohen, Cohen, West, & Aiken, 2003). The second criterion would be satisfied by the framework developed by Lee and See (2004; see Figure 3).

To this end, Bustamante (2009) conducted two experiments in which trust was subjectively measured after participants interacted with an 80% (FP or MP) reliable system for 30 minutes. Following this, a second 30 minute session was conducted in which reliance and compliance was measured using the same system. The system, in this case, varied between the two experiments. The first experiment utilized a simulated Engine Indicating and Crew Alerting System (EICAS) and the second experiment utilized the same secondary task as Rice (2009; see p. 15 of this work). Both systems issued advisories that indicated the engine/area was OK (engine status normal or no enemy tanks present) or an alarm which indicated that the system/area were not OK. When presented with an advisory the participant could ignore it, and continue with a primary task, or acknowledge it, and manually investigate the potential problem. Compliance was assessed by the number of times participants acknowledge the alarm and reliance consisted of the number of times participants ignored the OK advisory.

The results from both experiments were subjected to structural equation modeling (SEM) in which three models were proposed to derive the relationship between system characteristics (FP or MP) and participant dependence: partial mediation, full mediation, and no mediation. The results showed that from both experiments, the no mediation model consistently provided the best fit and all coefficients were statistically significant. From this, the author proposed that although dependence behaviors may be related to trust, it may be that trust is just a consequence of the system's characteristics (in this case FP versus MP).

As previously suggested, perceived reliability has been thought to determine trust in automation (e.g. Wickens & Hollands, 2000; Wiegmann et al., 2001). From this perspective, it would follow that trust should mediate the relationship between the system characteristic of reliability (which would be translated into perceived reliability) and the outcome behavior of operator dependence. Wiczorek and Manzey (2010) conducted a study that sought to further explore the surprising results reported by Bustamante (2009) by examining the relationship between perceived reliability, trust, and operator compliance. Using a multi-task scenario encompassing an alarm system of varying reliabilities (10% to 90% in 20% increments), participants were asked to rate their subjective trust in and perceived reliability of the system. The alarm system was FP only and, therefore, response rate was assessed by agreement rate with the alarm system when it indicated a problem needed to be fixed. Furthermore, the authors utilized a similar paradigm as Bustamante (2009) to satisfy the criterion of temporal precedence to determine causality.

From a simple mediator analysis, response rate seemed to depend on the perceived reliability of the system with no mediating effect of subjective trust in the alarm system. Furthermore, perceived reliability seemed to correspond to the actual reliability of the system where higher probabilities were significantly underestimated at the high end of the reliability spectrum and overestimated at the low end. However, there were no significant differences between the 30% and 50% conditions for actual and perceived reliabilities. These results could be explained by Kahneman and Tversky's (1983) prospect theory, which predicts that people often underestimate higher probabilities and overestimate lower probabilities, where the overlap between actual and perceived probability crosses over toward the lower end of the probability map. However, the authors did not interpret their results from this perspective.

Wiczorek and Manzey (2010) suggest that their results indicate that perceived reliability is what guides response rate. This would align well with the notion of probability matching suggested by some (e.g. Bliss, Gilson, & Deaton, 1995; Wiegmann et al., 2001). However, it may be misguided to suggest that perceived reliability and trust is the same thing. Again, the argument could be made that perceived reliability is too narrow to encompass the multidimensionality of trust. However, if perceived reliability does in fact determine the behavior observed, researchers could address the question from this perspective instead of assuming trust. If this perspective were adopted, it would simply be a matter of semantics and researchers would need to interpret situations more carefully as it may not be appropriate to assume that "over-trust" or "under-trust" is always the guiding psychological variable in situations denoted by unreliability.

However, both studies that reported a lack of evidence to suggest trust mediates the relationship between system characteristics and behavior failed to avoid the limitation of ignoring trust's multidimensionality and measured trust very narrowly. Bustamante (2009) limited trust rating responses to three items in which participants indicated the alarm system's perceived trustworthiness, dependability, and reliability. Wiczorek and Manzey (2010) limited responses to just one item indicating perceived trustworthiness. Both studies volunteered this as a limitation to their work and indicated that further research should incorporate a more multidimensional scale.

Although these studies suggest that trust may not mediate the relationship between system characteristics, such as perceived reliability and FP versus MP systems, and response behavior, the authors do not suggest that trust has no place. Instead, they highlight an intriguing notion that trust should not always be assumed to be the dominant psychological motivator of operator dependence. This implication would by no means negate the results from research that has inferred trust from behavior, but would modify the underlying assumptions from which this research proceeds. After all, what is often of utmost importance is to know how individuals behave, given the characteristics of a system. Furthermore, this research would also not undermine the importance of trust in automation research. Clearly, there could be conceivable instances in which an operator's trust would play a significant role in dependence behaviors.

Study Purpose

With the abovementioned perspectives concerning the role trust plays in operator dependence, the purpose of the current work would be to clarify the theoretical issues that may lead to interpretational inconsistencies. Specifically, the two main questions of

interest are, 'Does trust mediate the relationship between system characteristics and dependence and, if so, to what degree?' Investigating this question would clarify the amount of flexibility researchers have when inferring trust from behavior in signaling system paradigms. To address this question the current study will employ a similar multi-task paradigm as those of the aforementioned studies that includes a signaling system that have characteristics that vary in reliability (60% and 90%) and error bias (FP and MP). A Baseline (100%) control will also be implemented to aid comparisons. More specifically, participants will be asked to perform two tasks of the MATB II (resource management and the compensatory tracking tasks; Santiago-Espada, Myer, Latorella, & Comstock, 2011) while also responding to a signaling system that has the previously mentioned characteristics. Halfway through the task, participants will be asked to rate their trust in the alarm system. The second half of the task will be used to collect data concerning response behaviors. This would allow for a greater degree of certainty on the directionality of the effects of trust on behavior. Specifically, it would more clearly explain if trust may be the cause of response behaviors or if it is just a symptom of the characteristics of the system. To accomplish this, multiple simple mediation analyses will be conducted to test the relationship between subjective trust measures and dependence behaviors. The following section proposes multiple competing hypotheses based on the previously mentioned perspectives.

Trust Mediation Hypotheses

Generally, a variable is considered a mediator when it accounts for the relationship between the predictor and criterion (Baron & Kenny, 1986, p. 1176). For trust to act as a mediator between system characteristics and operator dependence two

assumptions must be met. First, as mentioned by Bustamante (2009), the logical ordering of variables should be established on theoretical and procedural grounds (i.e. trust must precede dependence temporally). More specifically, the outcome variable (i.e. dependence) should not cause the mediator (i.e. trust; Baron & Kenny, 1986; Preacher & Hayes, 2004). Second, there should be no measurement error (Baron & Kenny, 1986). This second assumption is much harder to meet because experimental psychology is tethered to the inescapable characteristic of depending on the measurement of the unobservable (Matthews, Davies, Westerman, & Stammers, 2008).

To illustrate the criteria of a mediator, the following section will elaborate on the diagrams presented in Figure 4. In Figure 4A, the relationship between X and Y is considered the total effect of X on Y, which is denoted as c. In Figure 4B, the relationship between X and Y is considered the direct effect of X on Y controlling for M (the mediator), which is denoted as c' to differentiate it from c in Figure 4A (Preacher & Hayes, 2004, p. 717). For a variable to be a mediator, Baron and Kenny (1986) propose the following: "(a) variations in levels of the independent variable significantly account for variations in the presumed mediator (i.e., Path a), (b) variations in the mediator significantly account for variation in the depended variable (i.e. Path b), and (c) when Paths a and b are controlled, a previously significant relation between the independent and dependent variables is no longer significant (c'), with the strongest demonstration of mediation occurring when Path c' is zero" (p. 1176). Preacher and Hayes (2004) relate this more succinctly as: "(1) X significantly predicts Y (i.e., $c \neq 0$), (2) X significantly predicts M (i.e. $a \neq 0$), and (3) M significantly predicts Y controlling for X (i.e. $b \neq 0$)"

(p. 717, parenthetical comments in original text). This is sometimes referred to as the

causal steps approach (Preacher & Hayes, 2004, 2008; Hayes, 2009).
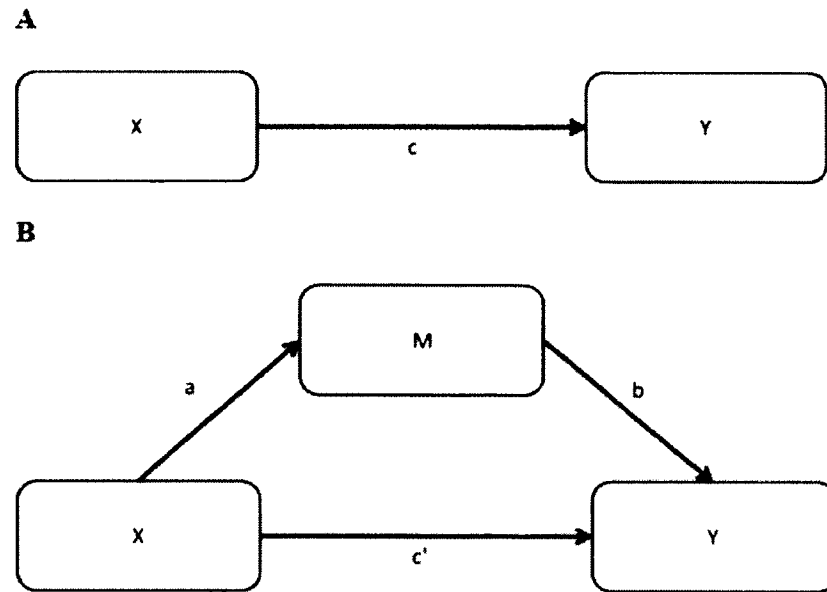
Figure 4. Graphical representation of a direct effect and simple mediation model. Model A illustrates a direct effect and model B illustrates a mediation design. Adapted from "*Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models,*" by K. J. Preacher and A. F. Hayes, Behavior Research Methods, 40 (3), p. 880. Copyright 2008 by Psychonomic Society, Inc.

However, to test if trust mediates the relationship between system characteristics and response behaviors, a technique advocated by Precher and Hayes (2004, 2008; Hayes, 2009) will be used. This method employs a version of the Sobel test to determine if the indirect effect of the mediator is significant. The traditional Sobel test assumes a normal distribution, which is not generally met in small sample sizes. Therefore, a non-parametric technique using bootstrapping is used which does not require such assumptions.

An additional, and important, comment is also warranted on the interpretation of indirect effects. Although the mediation effect assumes that path c was initially significant, an indirect effect does not. Therefore it is possible to find that an indirect effect is significant in the absence of a total effect (path c). According to the causal steps approach advocated by Barron and Kenny (1996), if there is not an initial significant effect of X on Y then further mediation analyses are not pursued. However, recently, some researchers have suggested that a significant indirect effect could be interpreted as a mediator (MacKinnon, Krull, & Lockwood, 2000; Hayes, 2009; Shrout & Bolger, 2002). For example, consider a model where a criterion variable exerts a significant effect on the outcome variable through two mediators working in opposite directions. This result would produce an outcome with a significant indirect effect absent of an initial total effect. This is because the two mediating variables, acting together, effectively cancel each other out in the estimation of the total effect. This scenario would appear as a non-significant total effect and, by the causal steps approach, would not lead to the conclusion that the indirect effect is acting as a mediator. Hayes (2009) has, therefore, suggested that the terminology does not affect the outcomes and failure to test for indirect effects in the absence of a total effect may lead researchers to miss instances in which X affects Y through unanticipated and/or potentially important mechanisms.

Finally, researchers generally assign verbal descriptors to the level of mediation, such as perfect, complete, or partial mediation. Partial mediation would suggest room for the presence of multiple mediators. Perfect or complete mediation, on the other hand, would not leave room for other mediators. Perfect mediation would be inferred if path c' (the direct effect) is zero once the mediator is controlled for. Barron and Kenny (1986)

caution, however, that perfect mediation is rare in psychology because of the presence of multiple mediators. Also, Preacher and Kelley (2009) suggest that because significance tests of the direct effect of X on Y are not independent of sample size, researchers may be "rewarded" for small sample sizes in which a non-significant direct effect is observed because of a lack of power. Therefore, although verbal descriptors will be used loosely in the current work, numerical effect sizes will be consulted to evaluate the degree of mediation if an indirect effect is observed in the data for the following hypotheses.

Hypothesis1a – Trust fully mediates the relationship between signaling system characteristics and response behavior (Dixon & Wickens, 2006; Dixon et al., 2007; Meyer, 2001; Rice, 2009). Meyer, (2001), Dixon et al. (2006), Dixon et al. (2007), and Rice (2009) did not explicitly state that trust is the only mediating factor present between system characteristics and response behavior. However, in its clearest theoretical form, inferring trust from behavior would implicitly assume that trust is the single most dominant mediator, which would allow for the methodology of omitting a subjective measurement. Thus, based on the causal steps approach, a hypothesis could be derived based on this assumption in which a significant indirect effect of X on Y through M is expected and path c' is also not statistically different from 0, when X is the system characteristic (reliability, FP, or MP), Y is the outcome behavior (reliance, compliance, signaling system agreement reaction time), and M is trust in the signaling system. In other words, if the independent variables (system characteristics) have no effect when the mediator (trust) is controlled, then perfect mediation can be assumed (Baron & Kenny, 1986). This would imply that trust fully mediates the relationship between system characteristics and it should be safe to infer trust from behavior with the present and

similar paradigms, as it would account entirely for the relationship between the predictor

and the outcome. Furthermore, mediation specific effect sizes should indicate that there is

little (or no) practical room left for the presence of other mediators in this relationship.

Hypothesis1b – Trust partially mediates the relationship between signaling

system characteristics and response behavior (Lee & Moray, 1992; 1994; Lee & See,

2004; Wiegmann et al., 2001). Some researchers have advocated caution when inferring

trust from behavior and have noted the importance of separating behavioral measures of

dependence from subjective measures of trust (Wiegmann et al., 2001). Indeed, from the

earlier review of the literature, there may be cases in which trust would not act as the

dominant mediator between system and behavior (Lee & Moray, 1994). However, this

research did not present results or suggest that trust does not play a role in this

relationship. To the contrary, this research would suggest that there is a relationship

between system characteristics, trust, and response behaviors. Therefore, from this

perspective it could be predicted that trust should significantly mediate the relationship

between system characteristics and behavior and decrease the relationship between X and

Y by some "non-trivial" amount. In other words, based on the causal steps approach, a

hypothesis could be derived where a significant indirect effect of X on Y through M is

observed and c' is reduced, where X is the system characteristic (reliability, FP, or MP),

Y is the outcome behavior (reliance, compliance, signaling system agreement reaction

time), and M is trust in the signaling system. Although this would offer support for the

theoretical models which assume trust as a prime mediator in this type of relationship, it

may also offer room to suggest that because trust mediates the relationship by a non-

trivial amount, inferences of trust from behavior may be safe to a certain degree in that it

at accounts for a significant portion of the variance. However, this would be a less certain assumption because it would suggest that the predictor is still influencing the outcome without the aid of the mediator via a single or multiple unchecked mediators in the relationship. Additionally, mediation specific effect sizes should indicate that there is some practical room left for the presence of additional mediators in this relationship.

Hypothesis1c – Trust does not mediate the relationship between signaling system characteristics and response behavior (Bustamante, 2009; Wiczorek & Manzey, 2010). As previously mentioned, a simple mediator analysis assumes that there is no measurement error. Measurement error in this case would tend to underestimate the effect of the mediator and overestimate the effect of the independent variable on the dependent variable when all of the coefficients are positive, which could cause a viable mediator to be overlooked (Baron & Kenny, p. 1177). The limitation mentioned by both Bustamante (2009) and Wiczorek and Manzey (2010) was that they used a narrow measure of trust. However, if these measures were adequately gauging operator trust in the signaling system, then a broader measurement tool should not overturn the results obtained in these studies. Thus, a final hypothesis could be derived based on this assumption where a significant indirect effect of X on Y through M is not observed, where X is the system characteristic (reliability, FP, or MP), Y is the outcome behavior (reliance, compliance, signaling system agreement reaction time), and M is trust in the signaling system. This would suggest that researchers should begin to interpret situations more carefully as it may not be appropriate to assume that "over-trust" or "under-trust" is always the guiding psychological variable in signaling system paradigms denoted by unreliability that is either MP or FP.

Trust Moderation Research Questions

In addition to trust acting as a mediator, trust may also act as a moderator. Barron and Kenny (1986) describe a moderator as "a qualitative (e.g. sex, race, class) or quantitative (e.g. level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable" (1174). To differentiate moderation from mediation, moderators always affect the outcome variable as independent variables, whereas mediators shift from effects to causes based on the analysis (Barron & Kenny, 1986). Conceptually, this relationship is represented in Figure 5A. From an analysis of variance (ANOVA) perspective, the statistical relationship can be represented more clearly as the interaction between an independent variable and a factor that identifies the degree to which the independent variable will operate. This statistical relationship is represented in Figure 5B as the interaction between an independent variable and moderator.
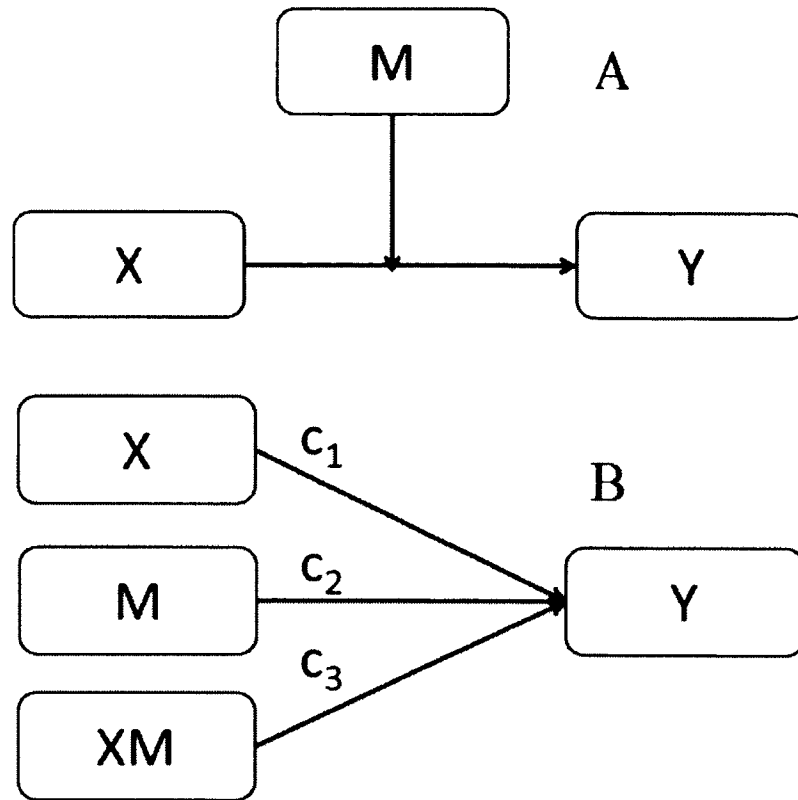
Figure 5. Conceptual model of moderation (model A) and statistical model of moderation (model B). Adapted from *"PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling,"* by A. F. Hayes. Retrieved from http://www.afhayes.com/public/process2012.pdf, p. 33. Copyright 2012 by Andrew F. Hayes.

Trust has been conceptualized as an enduring personality trait (e.g. Rotter, 1967; Gaines et al., 1997). In this way trust may act as a moderating variable and exert important effects on automation dependence behaviors, which may not necessarily be related to the characteristics of a specific automated system (Lee & See, 2004). Borrowing from Lee and See's (2004) framework, in this case a user's propensity to trust automation in general may inform each stage of the process outlined in Figure 3. Thus,

the manner in which a person evaluates his or her attitudinal trust of a particular system and subsequent response behavior could be differentially affected based on their propensity to trust similar technology in general. To test this relationship, moderation analyses were planned using a questionnaire developed to assess the participants' propensity to trust automation. If such a relationship exists after performing moderation analyses on the data, moderated mediation models would be used to explore this relationship further.

Behavioral Hypotheses

Generally, as automation reliability increases so does performance (Wickens & Dixon, 2007). However, FP and MP systems should differentially affect behavior in that false alarms more directly affect compliance and misses more directly impact reliance behaviors. Therefore, an interaction is expected in which the degree of the effect of reliability or response bias will depend on the error bias (Bliss et al., 1995; Meyer, 2001; Rice, 2009).

Main effects are also predicted for several other behavioral variables. Collapsing across signaling system error bias (MP, FP), higher reliabilities should yield greater performance in terms of participants being able to discriminate signal present versus signal absent (i.e. d') in response to signaling system advisories (Rice, 2009). Participants should agree more frequently with more reliable signaling systems (Bliss et al., 1995). Furthermore, higher reliabilities should yield quicker reaction times to signaling system advisories (Getty et al., 1995). Finally, participants should have higher signaling system scores when using more reliable systems (Bliss & Chancey, 2010). Multiple ANOVAs

should indicate that higher reliability rates yield better performance, more frequent agreement rates, and quicker reaction times for the secondary signaling system task.

Subjective Trust Hypothesis

It has been suggested that, generally, trust corresponds to the reliability of a system (Lee & See, 2004). Therefore, it is likely that participants will assign higher subjective trust ratings to more reliable signaling systems than to the less reliable signaling systems. However, whether or not trust then affects the subsequent dependence behaviors will be revealed by the simple mediation analysis.

METHOD

Design

The current study employed a 2 (false alarm prone or miss prone signaling

system) × 2 (90% or 60% reliable signaling system) split-plot design. An additional

baseline 100% reliable signaling system condition was also presented at the beginning of

each session. Signaling system error bias was a fixed, between-subjects variable with two

levels: 1) the signaling system's source of error was false alarms only – false alarm prone

(FP); 2) the signaling system's source of error were misses only – miss prone (MP).

Reliability was a fixed, within-subjects variable that indicated the probability that the

signaling system would correctly indicate a signal (system failure) or noise (system OK)

at two levels: 60% and 90%. The conditions were counterbalanced across participants

and presented in random order. Participants were informed of the reliability of the system

prior to each of the 60% and 90% experimental sessions. The 100% reliability condition

did not have an advertised reliability stated prior to the session so that a baseline response

rate could be obtained. Previous research has shown that advertising signaling system

reliability can have a significant effect on response behaviors independent of the actual

reliability (e.g. Bliss, Dunn, & Fuller, 1995).

Prior research has shown that explicitly informing participants of the reliability

level of an automated signaling system helps operators to more quickly and accurately

calibrate their level of trust to the reliability of the system over instances when reliability

levels are not explicitly stated (Bliss et al., 1995; Wang et al., 2009). Furthermore,

research has also demonstrated observed differences in subjective ratings of trust and

behavioral variables of interest in the current study within a relatively short experimental

session. For example, Wang et al. (2009) found differences in subjective trust and reliance using a similar signaling system paradigm where two levels of reliability and a control condition were manipulated within subjects in three 20-minute sessions. Bliss and Chancey (2010) found a main effect for reaction time in response to sensors that were associated with alarms systems that were 40% and 20% reliable. In that research, reliability was manipulated within subjects in two 30-minute experimental sessions.

After reviewing the literature, Wickens and Dixon (2007) concluded that signaling systems below a 71% reliability level were more detrimental to performance than if automation were not used at all. Based on these findings, the current study manipulated reliability levels to reflect a reliability level above the 71% cutoff (i.e. 90% reliable condition) and a reliability level below this cutoff (i.e. 60% reliable condition). This was done in an attempt to elicit measureable differences in the variables of interest.

Reliability of the signaling system was manipulated in a similar manner to Dixon et al. (2007), Bustamante (2009), and Rice (2009). The true state of the environment for each scenario contained 50% signal present situations (true system failures) and 50% signal absent situations (system is truly operational). The secondary signaling system task provided responses that corresponded to a signal detection response matrix. A hit indicates that the signaling system has correctly issued a system failure signal to a true system failure. A false alarm indicates that the system has issued a system failure signal in the absence of a true system failure. A miss indicates that the signaling system has issued an OK signal when a true system failure was present. A correct rejection indicates that the signaling system correctly issued an OK signal in the absence of a true system failure. Table 1 presents the breakdown of response rates by the system's error bias and

reliability according to the responses given by the signaling system in each experimental condition (see Appendix G for a breakdown of experimental script). Participants received feedback in regard to the appropriateness of their responses to these signals by auditory feedback from a voice that announced "Correct" or "Incorrect" after each response. In addition to the auditory feedback, participants were also presented with a "bank" of points that increased when the participant responded correctly and decreased when the participant responded incorrectly. Participants were not told what these points corresponded to, but were instructed to accumulate as many as possible.

Table 1

Breakdown of the detection response matrix for the false alarm prone (FP) and miss prone (MP) system according to reliability for the secondary signaling system task

	Baseline (100%)	*90% FP*	*60% FP*	*90% MP*	*60% MP*
Hits	30 (.50)	30 (.50)	30 (.50)	24 (.40)	6 (.10)
False Alarms	0 (.00)	6 (.10)	24 (.40)	0 (.00)	0 (.00)
Misses	0 (.00)	0 (.00)	0 (.00)	6 (.10)	24 (.40)
Correct Rejections	30 (.00)	24 (.40)	6 (.10)	30 (.50)	30 (.50)

Note. Numbers outside of parentheses represent the raw number of responses per category that occurred during the session. Numbers in parentheses represent the proportions of responses (out of the total number of responses during each session).

The paradigm employed in this study was intended to approximate the designs of the studies covered in the section "Behavioral Inferences of Trust" in that it required participants to perform multiple tasks at the same time (e.g. resource management, compensatory tracking, and signaling system response tasks), the task of primary interest in regards to trust was a simulated signaling system, and the independent variables manipulated were system reliability and error bias. This allowed for a more certain and appropriate conclusion when inferring trust from behavior utilizing this and similar designs.

To incorporate the recommendation by Bustamante (2009) and ensure temporal precedence, participant trust in the signaling system was obtained after they had interacted with it. Furthermore, for the simple mediation analyses behavioral measures of trust (reaction time and agreement rate) were considered only for analyses after this measure of trust had been taken. The proportion of errors made by the signaling system was equal for both the pre-trust administration and the post-trust administration so that an experimental artifact was not mistaken for a particular effect. The trust measure empirically developed by Jian et al. (2000), which gauges trust in automation, was used. This measure provides a broader assessment of trust than those employed in the studies previously mentioned; it does not limit trust to perceived reliance, confidence, or a single item of trustworthiness. Behavioral measures of trust suggested by Rice (2009) were employed to assess compliance and reliance. Compliance was operationally defined as the participants' agreement rate with the signaling system issuing a "System Failure" advisory and the reaction time to that agreement. Reliance was operationally defined as the participants' agreement rate with the signaling system issued a "System OK" advisory

and the reaction time to that agreement. Because this paradigm lends itself to signal detection analysis, participants' response bias (c) and sensitivity (d') were computed to represent overall performance among the conditions. Finally, similar to Bliss and Chancey (2010), a simple "alarm score" measure was implemented to indicate participant performance for correctly agreeing or disagreeing with signaling system advisories. The participant was provided a "bank" which contained an initial number of points. Each time the participant correctly agreed or disagreed with the signaling system advisory a point was added to the bank (i.e. agree with the system when it makes a hit or correct rejection or disagree with the system when it makes a false alarm or miss) and each time the participant incorrectly agreed or disagreed with the signaling system advisory a point was deducted (i.e. agree with the system when it makes a false alarm or miss or disagree with the system when it makes a hit or correct rejection). For this study the bank started at 20.

Participants

Forty-four participants (32 female, 12 male) participated in the 1.5-hour-long study for a five dollar gift card. Thirty-three of the participants were obtained by way of ODU's SONA system and received course credit for their participation. Eleven of the participants were recruited by email solicitation and did not receive a course credit equivalent. The mean age of participants was 26.89 years ($SD = 9.58$, min = 19, max = 28). All participants indicated having normal or corrected-to-normal visual acuity at the time of participation. None of the participants indicated having any color deficiency or hearing impairment. On average, participants indicated using a computer for work and recreation combined per week 31.93 hours ($SD = 28.32$; min = 4 hours, max = 140

hours). Participants indicated playing video/simulation games an average of 2.89 hours per week ($SD = 5.72$; min = 0 hours, max = 20 hours).

The study was approved by ODU's Institutional Review Board prior to the commencement of the project. An informed consent form was completed by all participants prior to beginning the experiment.

Materials/Apparatus

Informed Consent Form. Upon arrival, participants received and signed an Informed Consent Form (Appendix A). This form included a brief overview of the study as well as participation risks and benefits. The form also indicated that the participant could terminate the study process at any time.

Instruction sheet. Participants were provided an instruction sheet (Appendix C) after they completed the Informed Consent Form. This sheet contained standard information regarding the primary MATB II tasks and the secondary signaling system task. The experimenter read this sheet aloud to each participant and the participant was asked to read along using his or her own copy. Following this, all of the participants were given time to practice with the primary tasks and the secondary task.

Primary task – Multi-Attribute Task Battery (MATB II). The MATB II is a computer program that simulates the kinds of tasks that pilots perform during flight (Santiago-Espada et al., 2011). Participants were asked to respond to two of the tasks available in this battery: the compensatory tracking task and resource management task.

Compensatory Tracking Task. The purpose of this task is to simulate a key function pilots perform: maintaining level flight (see Figure 6). Using a joystick, participants attempted to keep a blue reticle as close to the center of a pair of crosshairs

as possible. The reticle, however, attempts to continuously drift in random directions. This is similar to a pilot flying an airplane while competing with environmental aspects such as wind. The root mean square deviation of the reticle from the center point was recorded from the joystick data to an output file every 15 seconds, reflecting stability of compensatory tracking performance.
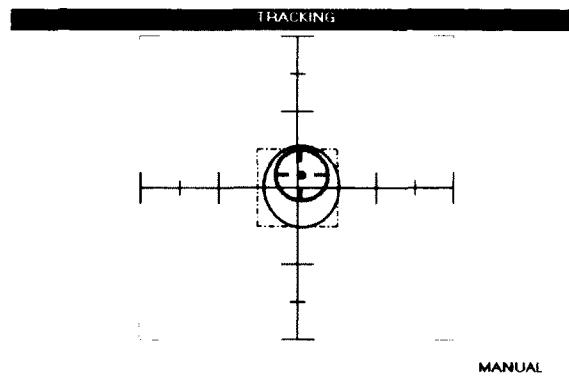
Figure 6. Screenshot of the compensatory tracking task.

Resource Management Task. This task is designed to represent a generalized fuel management system (see Figure 7). There are six tanks labeled A through F which contain fuel, represented by the green color inside each tank. Over time, fuel is consumed and the green "liquid" drops in tanks A, B, C, and D. Tanks E and F, however, are "supply" tanks and the fuel does not decrease in these tanks. The purpose of this task is to keep the fuel in tanks A and B at or above a specified level, which are indicated by the short blue dashes on both sides of each tank. Participants accomplished this task by transferring fuel from other tanks by separate pathways, which were represented by the blue lines connecting each tank. Each pathway had an associated pump which could be

activated or deactivated as needed. Pumps were represented by the squares along each

blue line, which were labeled 1 through 8 and have an arrow which indicated the

direction of the flow of the fuel. Participants turned pumps on or off by pressing the

corresponding number on a standalone wireless ten-key pad. A pump was indicated as

being on and transferring fuel from one tank to another if the pump was colored green,

depicted in Figure 7 by pumps 1, 2, and 5. A pump was not turned on or transferring fuel

if was white, depicted in Figure 7 by pumps 4, 7, and 8. Occasionally, however, a pump

would turn red and not transfer fuel or respond to the participant's input, which is

depicted in Figure 7 by pumps 3 and 6. Participants were told that this occurred from a

non-specified transient malfunction. Tank levels were recorded every 30 seconds.
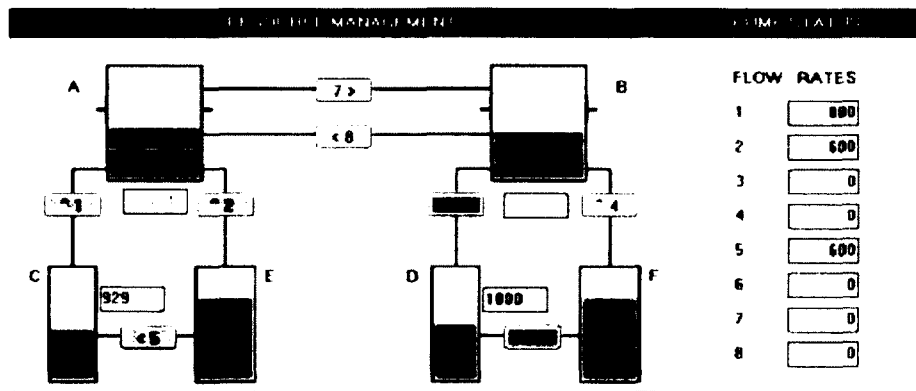
Figure 7. Screenshot of the resource management task.

Secondary task – Signaling system. Using SuperEdit 4.6™ the researcher

developed the secondary signaling system task, which was presented to participants using

SuperCard 4.6™ software hosted on a Macintosh desktop computer (see Figure 8).

Participants interacted with this task via mouse. This task was modeled after the system-

monitoring task of the MATB II. However there are several notable differences. The signaling system for this experiment was designed to issue an alert every 20 seconds to indicate the status of five gauges, which contained yellow indicators that fluctuate around the center of the display. As mentioned earlier, there were an equal number of times in which these gauges were within the normal range and outside of it. If the signaling system sensed that one of these indicators had deviated from the normal range, then it would issue a red "System Failure" alert, which was accompanied by the fire bell sound from a Boeing 757. If the signaling system did not sense that any of these gauges had deviated from the normal range, it would issue a green "System OK" signal, which was accompanied by a1000-Hz. tone. This system varied in its reliability or ability to accurately differentiate between signal present and signal absent (60% and 90%) and the types of errors it made (making false alarms or missing true alarms). During the initial experimental session this system was 100% reliable in its ability to differentiate between signal present and signal absent.
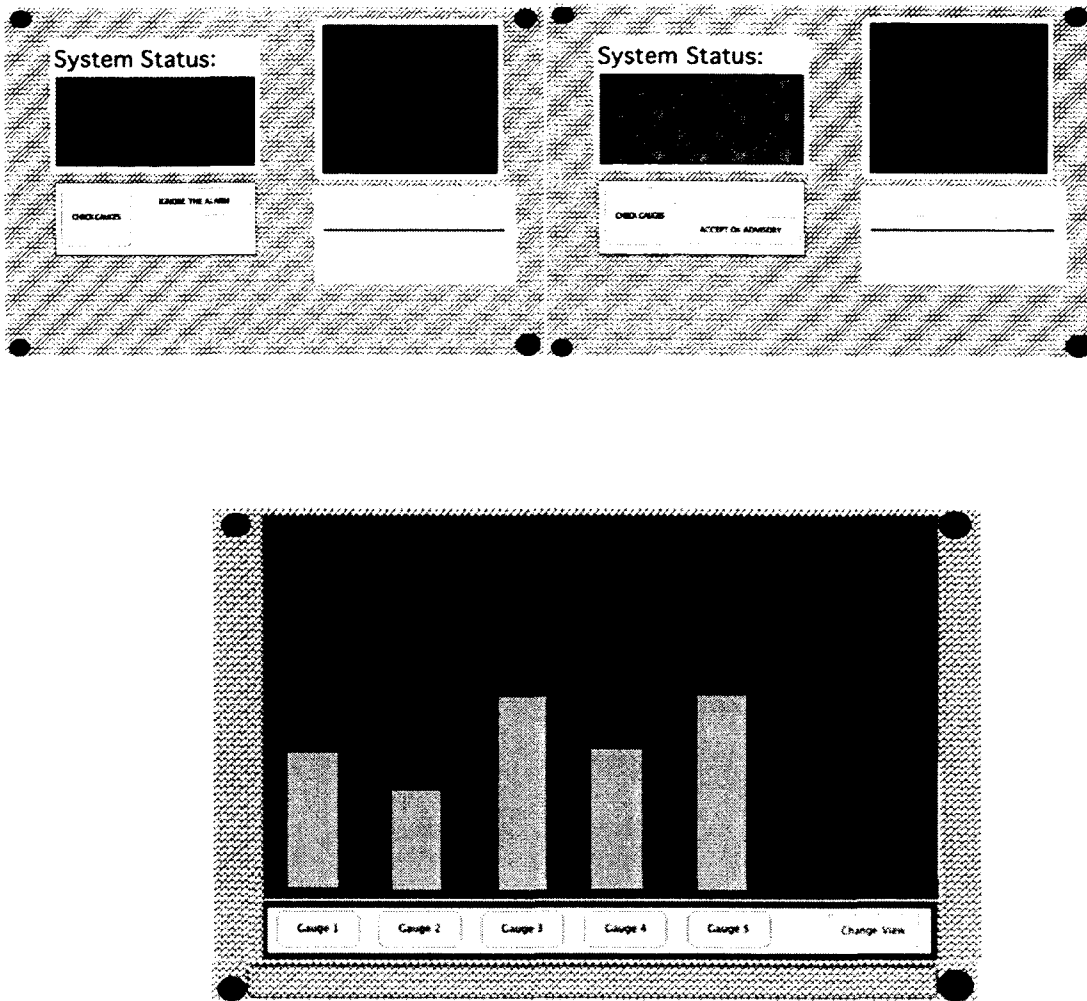
Figure 8. Screenshot of signaling system display during a system failure (top-left) and system OK (top-right). Screenshot of gauge display that appeared if participant selected the "Respond" button in response to a signal (bottom).

If the system issued a "System OK" signal, the participant could click the

"ACCEPT OK ADVISORY" button, indicating agreement with the alert. Alternatively,

the participant could click the "CHECK GAUGES" button, indicating disagreement with

the alert and the desire for further investigation. If this option was chosen, the participant

was presented with the five gauges. If one of the gauges was outside of the normal range,

then he or she would then click on that gauge to return it to normal and then continue working on the primary task. If, all gauges appeared normal, then he or she would then click on the "Change View" button, which would indicate that the participant needlessly investigated the potential problem. There was only one correct option on this display; the program audibly announced "Incorrect" if the participant failed to choose correctly. The participant could not progress until the correct option was chosen.

If the system issued a "System FAILURE" alert, the participant could click the "IGNORE THE ALARM" button, indicating lack of agreement with the alert and no need to visually inspect the gauges. Alternatively, the participant could click the "CHECK GAUGES" button, indicating agreement with the alert and desire for further investigation on the part of the participant. The participant was then presented with the five gauges as described in the previous paragraph.

The participant was provided with feedback about the accuracy of their decision. To facilitate this, participants were given auditory feedback and a bank which increased by 1 point for correct decisions and decreased by 1 point for incorrect decisions. If the "CHECK GAUGES" button was appropriately clicked, the program audibly announced "Correct" and the bank increased by 1 point. If this button was inappropriately selected, the program audibly announced "Incorrect" and the bank decreased by 1 point. If the "ACCEPT OK ADVISORY" or "IGNORE THE ALARM" buttons were appropriately clicked, indicating there truly was no gauge that needed to be fixed, the program audibly announced "Correct" and the bank increased by 1 point. If one of these buttons was inappropriately clicked, the program audibly announced "Incorrect" and the bank decreased by 1 point.

Trust Questionnaire. Jian et al.'s (2000) human-automation trust questionnaire was used to assess participant trust in the signaling system. The survey consisted of 11 items empirically developed to engage automation trust (Jian et al., 2000). An adequate internal-consistency reliability for this measure was observed, Cronbach alpha = .96. Furthermore, the scale has been used successfully in prior signaling system-based research (Gupta, Bisantz, & Singh 2002; Wang et al., 2009). Five of the items are statements that indicate operator mistrust toward the system and six of the items are statements that indicate operator trust toward the system. Each of the statements were accompanied by a 12 point Likert scale asking participants to indicate their agreement with the statement from "Not at all" to "Very much." Participants were asked to rate how much each item represented their attitude toward the signaling system (see Appendix D).

Dispositional Trust Questionnaire. Participants completed a Dispositional Trust Questionnaire prior to the experimental sessions. This was used for the moderation analyses. The questionnaire was adapted from the Jian et al. (2000) trust questionnaire and all of the items reflected the six trust statements of that questionnaire. The participant was asked to evaluate "The smoke alarm where I live," "The check engine gauge in most cars," "My alarm clock," "The computer I use most frequently," and "Modern technology in general (GPSs, computers, cell phones, vehicles, TVs, etc.)." An adequate internal-consistency reliability for this measure was observed, Cronbach alpha = .90 (see Appendix J).

Demographic Information Form. Participants completed a Demographic Information Form following the completion of the experimental sessions. Previous research has demonstrated that stereotype priming can affect subsequent attitudes and

performance (Shih, Pittinsky, & Ambady, 1999; Steele & Ambady, 2006). Because there

are differences associated with videogame use in regards to sex (Lucas & Sherry, 2004)

and age (Selwyn, Gorard, Furlong, & Madden, 2003) that may prime associated

stereotypes, the demographic questionnaire could have inadvertently affected variables of

interest. Therefore, participants completed the questionnaire (Appendix B) following the

collection of experimental data. This form asked participants to indicate their age, sex,

computer experience, video-game experience, hearing loss, visual acuity, and visual color

deficiency.

Opinion Questionnaire. After completing the Demographic Information Form,

participants filled out an Opinion Questionnaire (see Appendix E). This form recorded

the participant's interest and effort committed during the entirety of the experimental

session. Data from this questionnaire were used in the post-hoc correlational analyses for

exploratory purposes and to determine if experimental conditions were differentially

associated with effort and interest levels.

Reliability Questionnaire. Prior to each session a reliability questionnaire was

given as a manipulation check to ensure that participants understood each system's

characteristics (Appendix I). Participants indicated how often the signaling system should

be correct, how often the signaling system should be incorrect, and if the signaling

system were to make an error, which type of error it would be more likely to make.

Participants were required to get all three items correct before proceeding to the

experimental session. If a participant marked an item incorrectly in any of these sessions,

the researcher provided further instruction to help clarify why the item was marked

incorrectly. If the participant was still unable to understand why an item had been marked

incorrectly, the participant would have been excused from the experiment and his or her

data would have been excluded from the data analysis. No participant was excused for

this reason.

Apparati. The primary and secondary tasks were hosed on two separate desktop

computers (see Appendix F). Participants performed the MATB II tasks on an IBM

Pentium 4 CPU with a Windows XP operating system. The MATB II was viewed on a 12

inch IBM Think Vision monitor located on the right side of the desk. Participants

operated the compensatory tracking task with a Microsoft SideWinder Precision 2

Joystick and indicated their responses to the resource management tasks by pressing the

number keys (1-6) on a ten-key number pad. The SuperCard 4.6™ software which

presented the secondary Signaling System Task was hosed on a Macintosh desktop

computer, which utilized operating system X version 10.4.11. The Signaling System Task

was presented on a 12-inch Gateway FDP 1730 monitor located on the left side of the

desk. Participants indicated their alarm responses via mouse selection. The auditory

feedback and signaling system status updates were presented by RadioShack® PRO-100

Communications Headset headphones.

SPSS Mediation/Moderation Macro. To test the proposed mediation and

moderation models a Macro developed by Hayes (2012), referred to as PROCESS, was

used. The macro is compatible with SPSS and can be downloaded from

http://www.afhayes.com.

Procedure

Upon arrival at the testing location, participants were welcomed and asked to

complete the Informed Consent Form (Appendix A). Following this, participants were

asked if they had visual or hearing impairments. No participant indicated having any hearing issues, color deficiency, or impaired visual acuity without corrective devices. Participants were then given the Dispositional Trust Questionnaire (Appendix H).

Next, participants received instructions for the experimental sessions, which were also read aloud by the researcher (Appendix C). Instructions covered the basic premise and demands of the primary MATB II and secondary signaling system tasks. After hearing and reading these instructions, participants took time to practice all of the tasks alone and in combination. Once the researcher deemed that the participant was adequately familiarized with the experimental tasks, the first experimental session began.

Participants completed three 20-minute sessions with the primary and secondary tasks. All participants first interacted with a signaling system that was 100% reliable. Participants were not provided with any information indicating this system's reliability or error bias. To accommodate the within-subjects variable, following the interaction with 100% baseline, all participants were then presented with a signaling system that varied by reliability (90% and 60%). The remaining reliability level was presented in the second session. To accommodate the between-subjects variable, half of the participants were presented with signaling systems that were unreliable by making false alarms and the other half were presented with signaling systems that were unreliable by missing valid dangers. All conditions were counterbalanced.

For the 90% and 60% experimental conditions, participants were explicitly informed of the system's error characteristics prior to the session and then asked to fill out the Reliability Questionnaire (Appendix I) to ensure that he or she understood the characteristics of each system. Half way through all sessions (including the 100%

baseline) participants were presented with the trust questionnaire (Appendix D). Following this, participants completed the rest of the session, which comprised the behavioral data collection portion used for the simple mediation analyses.

Following the three 20-minute experimental sessions, participants then completed a background questionnaire (Appendix B) and an opinion questionnaire (Appendix E). Participants were then given a five-dollar gift card and awarded 2 SONA credits, which could be applied to course credit. They were then thanked for their participation, debriefed, and dismissed. The entire session lasted approximately 2 hours for each participant.

RESULTS

Following data coding, descriptive statistics were calculated and are presented in

Table 2. The data were inspected to ensure that there were no outliers, conditions had

equal numbers, and variables were generally normally distributed. Because the mediation

analyses used a non-parametric bootstrapping method, the assumption of normality was

not required. To identify outliers, an outlier-labeling rule with a multiplier of 2.2 was

consulted (Hoaglin & Iglewicz, 1987). Five outliers were identified in the 100% baseline

condition for reaction time upper limit. These data were brought within 1 ms above the

identified upper limit (6.15 seconds). One participant's data for the 100% baseline

condition were lost due to technical error. Therefore, those data were not included within

the repeated measures ANOVAs reported below. Leven's test for homogeneity of

variance were consulted to address this assumption for the between subjects measures.

Mauchley's tests were conducted to address the assumption of sphericity. If sphericity

was violated a Geisser-Greenhouse correction was used to interpret the data. An alpha

level of $p < .05$ was established to indicate statistical significance.

Table 2

Descriptive statistics

Source	*d'*	*c*	*Agree*	*RT*	*Trust*	*SS Score*
Baseline (100%)	3.03 (0.16)	-0.03 (0.03)	54.54 (0.94)	3.51 (0.19)	96.26 (4.74)	69.02 (1.87)
90%	2.34 (0.09)	0.09 (0.06)	55.81 (0.82)	2.70 (0.11)	85.65 (3.42)	62.14 (1.41)
60%	0.38 (0.05)	-0.10 (0.07)	38.33 (1.45)	2.83 (0.11)	40.14 (2.69)	27.91 (1.04)
90%	2.30 (0.09)	-0.09 (0.05)	55.71 (0.79)	2.72 (0.11)	85.27 (3.33)	61.86 (28.05)
60%	0.039 (0.05)	-0.09 (0.07)	39.43 (1.44)	2.86 (0.11)	40.25 (2.63)	28.05 (1.03)
FP	1.37 (0.08)	-0.31 (0.06)	48.21 (1.17)	2.81 (0.11)	66.32 (2.99)	4623 (1.21)
MP	1.32 (0.08)	0.13 (0.06)	45.93 (1.17)	2.72 (0.11)	59.21 (2.99)	43.68 (1.21)

Note. Descriptive statistics for repeated measures ANOVA (N = 43) (top). Descriptive statistics used for split-plot ANOVA (N=44) (bottom). Standard errors are in parentheses. Both instances of 90% and 60% reliability collapsed across error bias (FP, MP). c = response bias, d'=sensitivity, Agree = Agreement rate, RT = Reaction time (Secs), Trust = subjective estimates of trust in signaling system, SS Score = Signaling system score.

Simple mediation analyses using bootstrapping was conducted to test the competing trust mediation hypotheses. A moderation analysis was also conducted to explore the trust moderation research questions. Multiple 2 (FP, MP) × 2 (90%, 60%) split-plot ANOVAs were employed to test the behavioral and trust hypotheses. The 100%

baseline condition was omitted from these analyses. Multiple three-level, one-way, repeated measures ANOVAs were conducted on the reliability variable using the 100% baseline condition. Between-subjects comparisons were not made for error bias (FP and MP) against the 100% baseline conditions because of unequal sample sizes (i.e. baseline n = 43; FP n = 22; MP n = 22). Although ANOVA is generally robust to moderate violations of homogeneity of variance, this does not hold when sample sizes among the groups are not equal and even small violations can produce Type 1 error rates that are greater than .05 (Maxwell & Delaney, 2004, pp. 112-113).

Trust Mediation Analyses

Model 1 (X = Error Bias; Mediator = Trust; Y = Agreement Rate), Model 2 (X = Error Bias; Mediator = Trust; Y = RT), and Model 4 (X = Reliability; Mediator = Trust; Y = RT) did not yield a significant indirect effect, with all 95% CIs overlapping with 0 (see Figure 9 and Tables 3 and 4). This suggests that trust did not mediate the relationship between error bias or reliability and reaction time or between error bias and agreement rate.
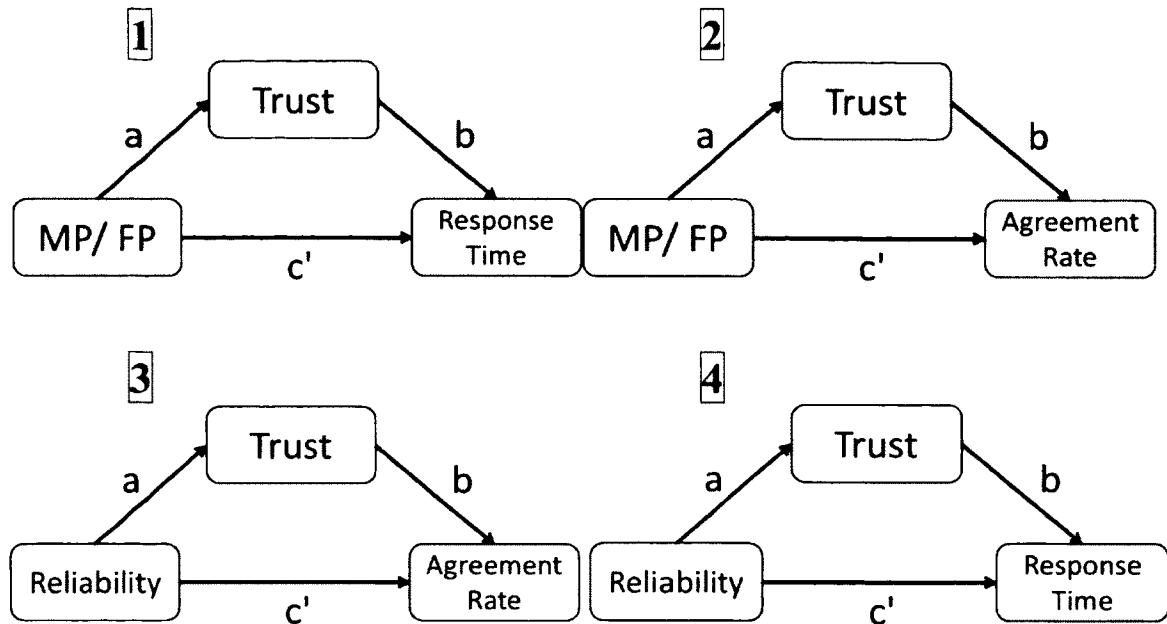
Figure 9. Models 1-4: Proposed models testing the simple mediation of signaling system characteristics (error bias and reliability) on behavioral indices of trust through subjective estimates of signaling system trust.

Model 3 (X = Reliability; Mediator = Trust; Y = Agreement Rate) yielded a significant indirect effect, where participants in the 90% reliability condition agreed with the signaling system 2.65 more times on average than those in the 60% reliability condition, as a result of the effect of their subjective trust in the signaling system (see Reliability-Agreement Rate in Table 4). Participants using the 90% reliable system indicated that it was 45.02 units more trustworthy (path a) than the 60% reliable system, and participants who felt the system was relatively more trustworthy agreed with the system more, path b = .06. Although the significant indirect effect of trust provides evidence that it is mediating the relationship between the reliability of the signaling system and propensity to agree, the direct effect (path c') is still significant (see Table 3).

This suggests that trust is a partial mediator in this relationship; however, the term,

"partial mediation" does not convey a numerical indication of the degree of mediation

and the statistical significance of c' is not independent of sample size (Preacher & Kelly,

2011).

Table 3

Path coefficients for the four trust mediation analyses

Source	Model 1	Model 2	Model 3	Model 4
a	7.11	7.11	45.02***	45.02***
	(6.42)	(6.42)	(4.27)	(4.27)
b	0.14***	-0.002	0.06**	-0.002
	(0.02)	(0.003)	(0.02)	(0.004)
c	0.73	0.03	9.18***	-0.11
	(1.29)	(0.17)	(0.84)	(0.17)
c′	-0.28	0.48	6.53***	-0.01
	(0.93)	(0.17)	(1.21)	(0.25)

$* p < .05; ** p < .01; *** p < .001$

Note: Standard errors are in parentheses. See Figure 9 above for corresponding model. Models tested do not include the baseline (100%) condition or the RT and agreement rates recorded prior to the administration of the trust questionnaire in the 90%, 60%, FP, and MP conditions.

Table 4

Indirect effects of trust for signaling system characteristics on behavioral responses through subjective estimates of signaling system trust.

Source	*Point Estimate*	*SE*	*Bootstrapping 95% CI*	
			Lower	**Upper**
MP/FP				
Agreement Rate	1.00	0.91	-0.78	2.77
RT	-0.02	0.03	-0.12	0.01
Reliability				
Agreement Rate	**2.65**	**0.95**	**1.13**	**4.93**
RT	-0.11	0.18	-0.18	0.24

Note: Data bootstrapped (5,000). Bolded numbers indicate a significant indirect effect. Models tested do not include the baseline (100%) condition or the RT and agreement rates recorded prior to the administration of the trust questionnaire in the 90%, 60%, FP, and MP conditions.

To help explore the extent to which trust mediated the relationship between signaling system reliability and response rate, two effect sizes are reported here. First, Preacher and Kelley (2011) proposed κ^2 as a measure of effect size for mediation analyses, where $\kappa^2 = 0$ implies that there is no linear indirect effect and $\kappa^2 = 1$ implies that the indirect effect is as large as it potentially could have been. The proportion of the maximum observed indirect effect was $\kappa^2 = .227$ (Bootstrapped: SE = .066, 95% CI [.104, .363]) for the mediating effect of trust between reliability and agreement rate. Second, the ratio of indirect to total effect of reliability on agreement rate was $P_m = .2883$ (Bootstrapped: SE = .106, 95% CI [.122, .550]). This measure of effect size is the most

frequently reported measure of mediation effect and is often interpreted as a proportion, where trust mediated approximately 28.83% of the total effect of reliability on agreement rate. However, caution should be taken when interpreting P_m in this manner as this value can be greater than 1 or less than 0 which would not make this a proportion. This measure is, however, unaffected by sample size and can facilitate inter-population comparisons, even when variables are not measured on the same scales across groups (Preacher and Kelley, 2011).

Trust Moderation Analyses

Participants' predisposition to trust technology did not moderate any relationships among the predictors (error bias and reliability) and outcome variables (agreement rate and RT). In other words, a significant interaction (path c_3) was not observed between the IVs and the moderator variable (measured by the Dispositional Trust Questionnaire, Cronbach's Alpha = .90) on the behavioral trust measures.
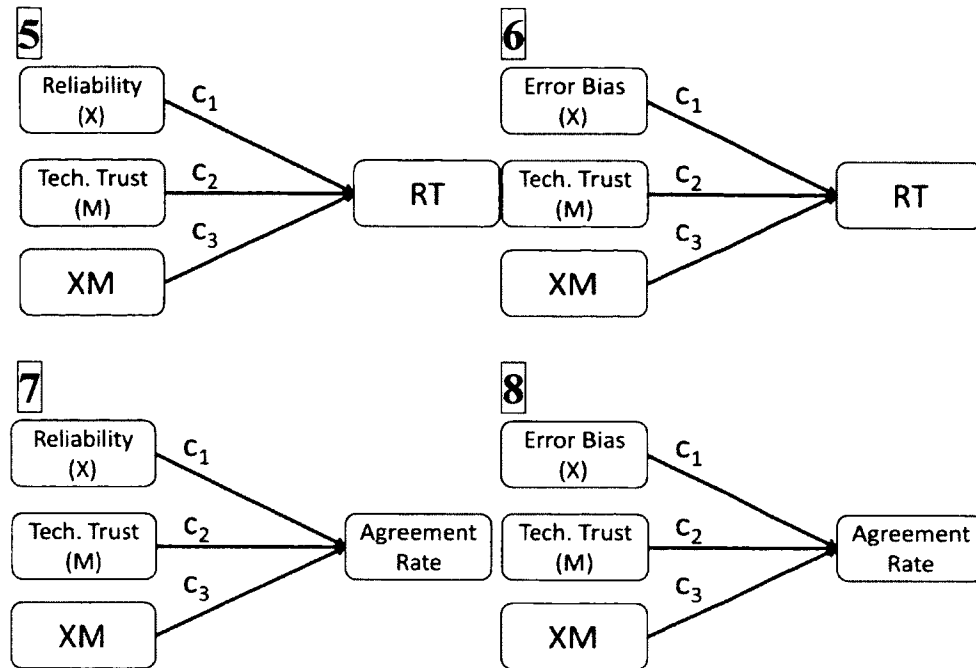
Figure 10. Models 5-8: Proposed models testing the moderation of signaling system characteristics (error bias and reliability) on behavioral indices of trust through predisposition to trust technology.

Table 5

Path coefficients for the four trust moderation analyses

Source	Model 5	Model 6	Model 7	Model 8
c_1	-0.1113 (-0.0005)	-0.0491 (0.1758)	9.1818*** (0.8415)	0.6713 1.3738
c_2	-0.0039 (0.0028)	-0.0045 (0.0031)	-0.0102 (0.0143)	-0.6713 (0.0241)
c_3	-0.0005 (0.0057)	0.0029 (0.0062)	0.0147 (0.0147)	-0.0314 (0.0483)

* p < .05; ** p < .01; *** p < .001

Note. Models tested do not include the baseline (100%) condition or the RT and agreement rates recorded prior to the administration of the trust questionnaire in the 90%, 60%, FP, and MP conditions. Standard errors are in parentheses. See Figure 10 above for corresponding model.

Primary Tasks

A split-plot ANOVA on compensatory tracking task performance did not reveal a significant interaction, $F(1, 42) = .787$, $p > .05$, observed power $= .140$, main effect of reliability, $F(1, 42) = .287$, $p > .05$, observed power $= .077$, or main effect of error bias, $F(1, 42) = 1.121$, $p > .05$, observed power $= .179$. A second split-plot ANOVA on resource management performance did not reveal a significant interaction, $F(1, 42) = 1.899$, $p > .05$, observed power $= .270$, main effect of reliability, $F(1, 42) = .752$, $p > .05$, observed power $= .135$, or main effect of error bias, $F(1, 42) = 2.629$, $p > .05$, observed

power = .354. These results suggest that the primary task performance was not differentially affected by the error characteristics of the secondary signaling system task.

A three-level, one-way, repeated measure ANOVA, with a Greenhouse-Geiser correction, revealed a significant main effect of reliability on compensatory tracking performance, $F(1.553, 66.800) = 10.776$, $p < .001$, $partial\ \eta^2 = .200$, observed power = .968. Follow-up Tukey tests indicated that the 100% baseline condition was associated with significantly less stable cursor movement ($M = 54.72$, $SE = 3.26$) than either the 60% reliability condition ($M = 49.01$, $SE = 3.11$) or the 90% reliability condition ($M = 49.52$, $SE = 2.82$), $F(1, 66.800) = 13.674 > F_{Tukey} = 5.78$. Similarly, a second three-level, one-way, repeated measure ANOVA revealed a significant main effect of reliability on resource management performance, $F(2, 86) = 20.536$, $p < .001$, $partial\ \eta^2 = .323$, observed power = 1.00. Follow-up Tukey tests indicated that the 100% baseline condition ($M = 2,140.25$, $SE = 102.03$) was associated with significantly worse fuel regulation performance than either the 60% reliability condition ($M = 2,463.19$, $SE = 101.26$), $F(1, 86) = 34.798 > F_{Tukey} = 5.78$, or the 90% reliability condition ($M = 2,512.40$, $SE = 92.21$), $F(1, 86) = 26.20 > F_{Tukey} = 5.78$. Because the 100% baseline condition was always presented first, these results suggest that participants were acclimating to these tasks.

Sensitivity

A split-plot ANOVA revealed a significant main effect of reliability on participant sensitivity, $F(1, 42) = 341.074$, $p < .000$, $partial\ \eta^2 = .890$, observed power = 1.00, where participants in the 90% condition ($M = 2.30$, $SE = 0.09$) were significantly more sensitive than those in the 60% reliable condition ($M = 0.039$, $SE = 0.05$). However, a significant interaction between reliability and error bias, $F(1, 42) = 4.456$, $p = .041$,

partial η^2 = .096, observed power = .451, indicated that the difference in sensitivity between the reliability conditions was more pronounced for the FP systems than the MP systems. The main effect of error bias on sensitivity was not significant, $F(1, 42) = .257$, $p > .05$, observed power = .079.

A Mauchley's test indicated that the assumption of sphericity was not met for comparisons across reliability levels, $\chi^2(2) = 11.994, p = .003$. A three-level, one way, repeated measure ANOVA, with a Greenhouse-Geisser correction, revealed a significant main effect of reliability, $F(1.597, 67.054) = 163.996$, $p < .001$, *partial* $\eta^2 = .796$, observed power = 1.00. Follow-up Tukey tests indicated that participants showed significantly more sensitivity in the 100% baseline condition ($M = 3.03$, $SE = 0.16$) than in the 90% reliability condition ($M = 2.34$, $SE = 0.09$), $F(1, 67.054) = 17.265 > F_{Tukey} =$ 5.78, or the 60% reliability condition ($M = 0.038$, $SE = 0.05$), $F(1, 67.054) = 244.134 >$ $F_{Tukey} = 5.78$. Finally, participants showed more sensitivity in the 90% reliability condition than in the 60% reliability condition, $F(1, 67.054) = 131.552 > F_{Tukey} = 5.78$.

Response Bias

A split-plot ANOVA did not reveal a significant interaction for reliability and error bias on response bias, $F(1, 42) = 1.633$, $p > .05$, observed power = .239. However, a significant main effect of error bias on response bias was observed, $F(1, 42) = 26.61$, p $< .000$, *partial* $\eta^2 = .388$, observed power = .999, indicating that participants in the FP group ($M = -0.31$, $SE = 0.06$) had more liberal response bias than the MP group ($M =$ 0.13, $SE = 0.06$). In other words, participants in the FP group were more likely to respond as though there was a signal present (i.e. gauge needing to be fixed) independent of their

actual performance sensitivity (d'). A significant main effect of reliability on response bias was not observed, $F(1, 42) = .005$, $p > .05$, observed power = .051.

A three-level, one-way, repeated measure ANOVA did not reveal a significant main effect of reliability on response bias, $F(2, 42) = .929$, $p > .05$, observed power = .059.

Agreement Rate

A split-plot ANOVA revealed a significant main effect of reliability on participant agreement rate, $F(1, 42) = 113.219$, $p < .001$, *partial* $\eta^2 = .729$, observed power = 1.00, where participants agreed with the signaling system significantly more in the 90% condition ($M = 55.71$, $SE = 0.79$) than in the 60% reliable condition($M = 39.43$, $SE = 1.44$). Neither a significant interaction for error bias and reliability on response rate, $F(1, 42) = .095$, $p > .05$, observed power = .060, or main effect of error bias on response rate, $F(1, 42) = 1.889$, $p > .05$, observed power = .269, occurred.

A Mauchley's test indicated that the assumption of sphericity was not met, $\chi^2(2) = 12.590$, $p = .002$, for comparisons across reliability levels. A three-level, one-way, repeated measure ANOVA, with a Greenhouse-Geisser correction, revealed a main effect of reliability, $F(1.582, 66.434) = 88.332$, $p < .001$, *partial* $\eta^2 = .678$ observed power = 1.00. Follow-up Tukey tests indicated that participants agreed significantly more often in the 100% baseline condition ($M = 54.54$, $SE = 0.94$) than in the 60% reliability condition ($M = 38.33$, $SE = 1.45$), $F(1, 66.434) = 96.568 > F_{Tukey} = 5.78$. Participants agreed more often in the 90% baseline condition ($M = 55.81$, $SE = 0.82$) than in the 60% reliability condition ($M = 38.33$, $SE = 1.45$), $F(1, 66.434) = 112.41 > F_{Tukey} = 5.78$.

Reaction Time

A split-plot ANOVA revealed a significant main effect of reliability on reaction time, $F(1, 42) = 4.360$, $p = .043$, *partial* $\eta^2 = .094$, observed power = .532, where participants in the 90% condition ($M = 2.72$ secs, $SE = 0.11$) reacted significantly quicker when agreeing with the system advisory than in the 60% reliable condition ($M = 2.86$, $SE = 0.11$). The analysis did not reveal a significant interaction, $F(1, 42) = .595$, $p > .05$, observed power = .117, or main effect of error bias on reaction time, $F(1, 42) = 0.18$, $p > = .05$, observed power = .052.

A Mauchley's test indicated that the assumption of sphericity was not met for comparisons across reliability levels, $\chi^2(2) = 12.590$, $p = .002$. A three-level, one-way, repeated measure ANOVA, with a Greenhouse-Geisser correction, revealed a main effect of reliability, $F(1.403, 58.941) = 29.839$, $p < .000$, *partial* $\eta^2 = .415$, observed power = 1.00. Follow-up Tukey tests did not indicate any differences among the levels. A non-planned contrast assessing participant reaction time in the 100% baseline ($M = 3.51$ secs, $SE = 0.19$) against the collapsed means of the 60% and 90% reliability conditions ($\psi = 2.77$ secs.) was significant using a Scheffe correction, $F(1, 42) = 35.609$, $p < .001$, *partial* $\eta^2 = .459$, observed power = 1.00 ($F=35.609 > F_{Scheffe} = 6.46$), indicating that when participants were in the initial 100% baseline condition they took longer to agree with the signaling system than when they were in the 90% and 60% reliability conditions.

Signaling System Task Score

A split-plot ANOVA revealed a significant main effect of reliability on signaling system score, $F(1, 42) = 387.553$, $p < .001$, *partial* $\eta^2 = .902$, observed power = 1.00, where participants in the 90% condition ($M = 61.86$, $SE = 28.05$) scored significantly

higher than when in the 60% reliable condition ($M = 28.05$, $SE = 1.03$). The analysis did not reveal a significant interaction, $F(1, 42) = 2.042$, $p > .05$, observed power $= .287$, or main effect of error bias, $F(1, 42) = 2.230$, $p > .05$, observed power $= .309$.

A three-level, one-way, repeated measure ANOVA revealed a significant main effect of reliability, $F(2, 84) = 255.517$, $p < .000$, *partial* $\eta^2 = .859$, observed power $= 1.00$. Follow-up Tukey tests indicated that the 100% baseline ($M = 69.02$, $SE = 1.87$) was different from the 90% reliability group ($M = 62.14$, SE $= 1.41$), $F(1, 84) = 12.481 >$ $F_{Tukey} = 5.78$. The 100% baseline condition ($M = 69.02$, $SE = 1.87$) was significantly different from the 60% reliability condition ($M = 27.91$, $SE = 1.04$), F(1, 84) $= 445.349 >$ $F_{Tukey} = 5.78$. The 90% reliability condition ($M = 62.14$, $SE = 1.41$) was significantly different from the 60% reliability condition ($M = 27.91$, $SE = 1.04$), F(1, 84) $= 308.723 >$ $F_{Tukey} = 5.78$.

Signaling System Trust

A split-plot ANOVA revealed a significant main effect of reliability on subjective trust, $F(1, 42) = 105.607$, $p < .001$, *partial* $\eta^2 = .715$, observed power $= 1.00$, where participants in the 90% reliability condition ($M = 85.27$, $SE = 3.33$) rated the signaling system significantly higher than in the 60% reliable condition ($M = 40.25$, $SE = 2.63$). Although not significant by conventional levels, FP systems ($M = 66.32$, $SE = 2.99$) were rated as being more trustworthy than MP systems ($M = 59.21$, $SE = 2.99$), $F(1, 42) =$ 105.607, $p = .090$, *partial* $\eta^2 = .067$, observed power $= .396$. A significant interaction was not observed between reliability and error bias for subjective trust, $F(1, 42) = .204$, $p >$ .05, observed power $= .073$.

A three-level, one-way, repeated measure ANOVA revealed a significant main effect of reliability, $F(2, 84) = 71.833$, $p < .001$, *partial* $\eta^2 = .631$, observed power = 1.00. Follow-up Tukey tests indicated that the 100% baseline condition ($M = 96.26$, $SE = 4.74$) showed higher ratings of trust than the 60% reliability condition ($M = 40.14$, $SE = 2.69$), $F(1, 84) = 127.254 > F_{Tukey} = 5.78$. The 90% reliability condition ($M = 85.65$, $SE = 3.42$) had higher ratings of trust than the 60% reliability condition ($M = 40.14$, $SE = 2.69$), $F(1, 84) = 127.254 > F_{Tukey} = 5.78$.

Correlational Analyses

Pearson correlations were used to test for relationships between subjective and behavior based measures of trust (see Table 6). Because making a Type II error was a greater concern than making a Type I error in this instance, no alpha corrections were made for these analyses (Wickens, 1998). Several of the correlations were significant. As would be expected, higher agreement rates were associated with higher subjective ratings of the system for both the 100% baseline condition and 60% conditions. Interestingly, higher agreement rates in the 90% condition significantly correlated with higher subjective estimates of trust in the 100% baseline condition but not the 90% condition. This could have occurred because both systems were highly reliable. Although agreement rate and subjective trust seemed to correspond, reaction time did not correlate with subjective estimates of trust.

Table 6

Correlations between objective and subjective measures of trust

	Sub. BL	Sub. 90	Sub. 60	RT BL	RT 90	RT60	Agr. BL	Agr.90	Agr 60
Sub. BL	.	.372*	-.114	-.100	-.032	.034	.806**	.319*	-.044
Sub. 90		.	-.035	.004	-.019	-.104	.265	.271	.077
Sub. 60			.	.053	-.057	-.072	-.081	.054	.319*
RT BL				.	.796**	.691**	-.223	.187	.080
RT 90					.	.826**	-.185	-.109	.033
RT 60						.	-.118	.040	.160
Agr. BL							.	.327*	.086
Agr. 90								.	.046
Agr. 60									.

* < .05; **< .01

Note. No alpha corrections made.

DISCUSSION

The effect reliability had on agreement rate aligned well with the partial

mediation hypothesis (Lee & Moray, 1992; 1994; Lee & See, 2004; Wiegmann et al.,

2001). Trust exerted a significant indirect effect on whether or not participants decided to

agree with system suggestions above and beyond the initial total effect of reliability. The

initial direct effect of reliability, however, was still significant after controlling for trust.

Conversely, indirect effects of trust on behavior were not observed between the

error bias of the signaling system for either agreement rate or reaction time. Furthermore,

trust did not seem to mediate the relationship between system reliability and reaction

time. These results align with the hypothesis suggesting that trust does not mediate the

relationship between system characteristics and response behaviors (Bustamante, 2009;

Wiczorek & Manzey, 2010).

Until recently, research has generally supported the position that trust plays a

large role in mediating signaling system interactions. The mediation analyses reported in

the current study suggest that caution be exercised when inferring trust as the prime or

sole mediator affecting response behaviors. The following sections discuss the

implications of the results reported for the mediation, moderation, and performance

analyses.

Consideration of Other Meditating Variables

It is important to recall that although Rice (2009) operationally defined trust as a

behavior, he did not suggest that the psychological variable of trust and the behavioral

variable of dependence always perfectly correlate. In applied setting, dissociations

between trust and behavior could arise due to legal obligations to respond to system

advisories regardless of personal trust, such as in aviation (Bliss & Gilson, 1998).

Additionally, dissociations could also arise in situations characterized by high workload,

where the operator may intend to intervene but be too busy to respond. It is unclear if

differential levels of workload were affecting participant behavior in this way for the

current study. Differences in primary task performance among the conditions (excluding

the 100% baseline) were not observed. There were, however, differences in secondary

task performance: participants discriminated signal present from signal absent advisories

much better and also had higher signaling system scores with a more reliable system. It

may have been that participants chose to focus more of their attention on the primary

tasks when the reliability of the system was low to maintain consistent performance

levels (e.g. Bliss & Dunn, 2000; Bliss 2003). This interpretation, in the context of this

study, may be less likely because participants received no feedback about the validity of

individual advisories. However, because participants were aware of error characteristics

they may have adopted a particular performance strategy prior to beginning the tasks

(Bliss, 2003). A workload questionnaire may have been useful for clarifying how

workload interacted with trust to affect the outcome behaviors with this particularly

opaque system. The current analyses indicated that there was significant room left for

intervening mediators. Future research should seriously consider the presence of

workload on agreement rate and reaction time in place of the assumption that system trust

is largely what is motivating the behavior (cf. Bliss & Dunn, 2000).

A second important consideration suggested by Rice (2009) is that the option or

availability to validate the raw data against a signaling system's recommendation can

lead to very different behavioral outcomes. To illustrate this point, he offered the

example of a fire alarm sounding in a building. Individuals who can see or smell the fire

may act very differently than those who cannot perceptually verify the presence of the

fire and only hear the alarm. Similarly, if the operator has the option to verify the raw

data against the signaling system's recommendation, dependence behaviors are more

predictable than when the operator must blindly accept or reject its recommendation

(Bliss, 2003; Rice, 2009; Sorkin & Woods, 1985). Consequently, if knowledge of the

system error characteristics is not available, behaviors exhibited during interactions with

highly opaque systems, or that does not offer the option of validating the system

recommendations by referencing the raw data, may become unpredictable. Lee and See

(2004) suggest that trust in the automation also depends on the information available.

Therefore, although participants were made explicitly aware of the systems' error

characteristics, the reason that trust did not consistently act as a strong mediator may

have depended on the relative opacity of the system.

Extending this line of thought, Bustamante (2009) and Wiczorek and Manzey

(2010) also used opaque systems. The ability to evaluate individual recommendations

could play an important role in affecting how trust mediates the relationship between the

system's error characteristics and the outcome behaviors. In other words, if the system's

recommendations cannot be evaluated, then trust may play less of a role because the

operator is incapable of determining *why* or *how* the recommendation was reached. The

operator, then, may either choose to agree or disagree with a system based on prior

interactions and probabilities of errors or may helplessly depend on the system and agree

with it every time. In either situation, operator trust may fluctuate independent of the

response behavior or interact with other mediating variables in more complex ways. For

example, Bliss (2003) noted that when participants did not have access to raw data they tended to have extreme response patterns (i.e. respond to all recommendations or respond to none). He suggested that this tactic is adopted to conserve attentional resources. Yet, an operator's propensity to trust in general or his or her trust of similar systems may still play a role in forming behaviors (see dispositional trust discussion below).

Lee and See (2004) suggest that designing trustable automation should consider incorporating systems that show the processes and algorithms involved. The collective results from the current study, Bustamante (2009), and Wiczorek and Manzey (2010), would provide supporting evidence for this recommendation. It may be that system opacity plays an important role in governing the degree to which trust plays a role at all. One of the problems associated with opaque systems is that they can be related to operators having deficient mental models of how the system is deriving recommendations. These deficiencies can be the result of inadequate training, the relative complexity of the system, or both (Billings, 1996). Operators often have deficient mental models of complex systems. In such situations, they may totally rely on automation to act or to inform. Consequently, trust may be playing less of a role or interacting with other variables to produce the behavior.

If the signaling system is relatively opaque, researchers should consider the role of other mediating variables in determining response behaviors other than trust, such as operator self-confidence (Lee & Moray, 1994), effort to engage the automation (Kirlik, 1993), workload (Rice, 2009), and possibly situational understanding. Designers should consider creating less opaque systems that provide details for how the signaling system develops recommendations. A potential problem with this approach, however, would be

that it could cause the operator to divert attentional resources to system information. This could then create a situation characterized by high workload and reduced situation awareness. To circumvent this problem, a second design solution may be to implement likelihood alarm displays that issue different types of signals based on the probability that there is an actual underlying problem (Sorkin, Kantowitz, & Kantowitz, 1998). Preliminary research has shown that this could be an effective solution (Bustamante, 2008). Future research should address the issue of signaling system opacity to determine its function in relation to subjective estimates of trust and operator behavior.

Although one of Bustamante's (2009) secondary signaling system tasks was opaque, within his analyses he also incorporated an additional secondary signaling system task that was slightly more transparent. Bustamante (2009) had participants complete a primary tracking and system-monitoring task in addition to the same secondary signaling system task used by Rice (2009). Rice (2009), alternatively, had participants complete this task exclusively and view images from a chin rest to control for head movement. This would suggest that although both iterations of the signaling system task would be considered relatively transparent, because of the introduction of a dual task scenario along with the increased potential for head movement to access the raw data, Bustamante's (2009) version could have introduced more workload into the task. Again, the presence of workload may have suppressed trust as a mediator. Therefore, it is difficult to determine if Rice (2009) would have identified trust as mediating the relationship between error bias and agreement rate.

Trust did not mediate the relationship between any of the signaling system characteristics and reaction time in the current study. These results align well with

Wiegmann et al.'s (2001) findings that reaction time did not correlate with confidence in and perceived reliability of the signaling system (used as indices of trust). Inferring trust from reaction time, therefore, should be done cautiously. For example, Bliss and Chancey (under review) compared sensor reaction training strategies for a 20% and 40% reliable alarm systems in two experiments. For both experiments, there were no observed differences for alarm reaction times among the conditions but participants did trust the 40% reliable systems more than the 20% reliable systems. The current study found a small (although significant) difference between the higher and lower reliability conditions (excluding the 100% baseline). However, this difference was in the order of 140ms. The results of this study do not provide sufficient evidence to suggest that trust was playing any major role in affecting this difference.

The Role of Dispositional Trust on Response Behaviors

The current study did not show that participants' dispositional trust affected their response behaviors. It is unclear, however, if these results were observed because dispositional trust does not play a role in how participants behaved or because of methodological issues. Although the questionnaire for dispositional trust used here showed adequate internal consistency, further psychometric evaluation should be conducted to ascertain the validity of this tool as a sensitive measure of dispositional automation trust. Second, the current investigation lacked sufficient statistical power to test both moderation and mediation. The ability to evaluate operator dispositional automation trust would be valuable and could lead to more appropriate matches of human-automation teaming and improved individual operator training.

Performance and Subjective Trust Outcomes

Although trust did not seem to mediate most of the tested relationships, specific

behavioral and trust hypotheses were generally upheld. This would indicate that the

paradigm employed in the current study produced similar effects on response behaviors

and subjective estimates of trust observed in other studies (e.g. Chancey & Bliss, 2012;

Lee & See, 2004; Wang et al., 2009; Wiegmann et al., 2001). Because the characteristics

of the 100% reliability system were not advertised, the results that included this condition

must be carefully interpreted. Advertising reliability can play a significant role in

motivating subsequent behavior and trust, independent of actual reliability (e.g. Bliss et

al., 1995; Chancey & Bliss, 2012; Wang et al., 2009). These results can, however,

provide insight into the effectiveness of the 100% baseline condition serving as an

additional and adequate familiarization session.

Error bias seemed to have little effect on the dependent variables observed, except

for response bias. Participants in the FP conditions were more likely to respond as though

there was a signal present (i.e. gauge needing to be fixed) independent of their actual

performance sensitivity (d') than those in the MP conditions. These results align well with

Meyer's (2002) findings that the type of error the system commits differentially affects

response bias.

Reliability had the most pronounced effect on the dependent measures.

Participants showed higher sensitivity scores for the more reliable systems and were

generally better able to discriminate when the system was making accurate or inaccurate

recommendations. Participants also accumulated more points by correctly agreeing and

disagreeing with the system when it was more reliable. The higher scores were due to the

high agreement rates observed for higher reliable systems, where participants tended to agree with the recommendation of the system based on its reliability. These results align well with Wiegmann et al.'s (2001) suggestion that users tend to calibrate their agreement rate to the reliability of the system or probability match (e.g. Bliss et al. 1995), even at the expense of better performance outcomes by simply agreeing with the aid every time. Together, these results suggest that reliability played a large role in determining the overall performance for interacting with the signaling system.

Reaction time was significantly quicker for the 90% over the 60% reliability conditions without the 100% baseline condition. This difference, although significant, was small (140ms). The degree, to which this difference is practically significant, rather than statistically significant, should be considered in relation to the context in which it is applied (Wickens, 1998). For example, this difference may be of practical significance for some tasks in aviation, nuclear power, and medicine. In the context of the mediation analysis, it is unclear how reliability could have been causing this difference, as trust did not seem to be a factor. Differences in reaction time between the 90% and 60% conditions, however, were washed out with the inclusion of the 100% baseline condition. A contrast comparing the combination of the 90% and 60% condition against the 100% baseline revealed that participants were much slower in the initial 100% baseline condition. This result is likely due to an order effect present in the study. The 100% baseline condition was always given before the other two reliability conditions and the observed differences could be attributed to additional familiarization taking place as participants attempted to coordinate all three tasks simultaneously. Similar differences in

the primary task results also indicated that performance levels were significantly lower in the 100% baseline condition than in the 90% and 60% reliability conditions.

Participants in higher reliability conditions assigned higher levels of trust to the signaling system. These results provide additional evidence to suggest more reliable aids are generally also trusted more (Lee & See, 2004). The mediation analysis, however, did not suggest that trust exclusively caused participants to then agree more often with the aid recommendations. As previously mentioned, additional mediating variables not accounted for were likely causing participants to agree with the recommendations based on the reliability of the system. Therefore, although trust, agreement rate, and reaction time each changed in predictable directions based on reliability, in the case of reaction time trust played no role in this incremental movement and in the case of agreement rate trust was only partially accounting for this movement. However, it is important to consider the validity of the trust questionnaire used in the current study. Although the automation-trust questionnaire developed by Jian et al. (2000) has been shown to have convergent validity as a tool for evaluating customers' experiences during automated interactions (Safar & Turner, 2005), it is unclear whether differences were not observed because of inappropriate application. Further research should replicate these findings using a validated human-automation trust questionnaire tailored for signaling systems.

Conclusion

Although trust in automation may be considered a primary influence on operator behavior, the results of this study do not strongly support this conclusion. Because operator trust has been assumed to play such a large role in determining behavior, other variables in signaling system paradigms have been frequently overlooked. Parasuraman

and Riley (1994) suggest other intervening variables, such as situations characterized by high levels workload, may determine dependence behaviors even if the automation is generally reliable and trustworthy. Furthermore, Parasuraman and Riley (1994), suggest that behavior may be determined by a complex interaction of multiple variables besides trust, such as those modeled by Riley (1994). As previously mentioned, an operator's self-confidence may play an integral role in his or her decision to depend on an automated teammate or execute tasks manually (Lee & Moray, 1994). Apart from cognitive factors, the operator response to signals may be mandated, which occurs in aviation (Bliss & Gilson, 1998). Researchers have advocated training operators to appropriately calibrate their level of trust to the actual reliability of the system (Lee & See, 2004). However, in light of the findings of this study it may be more effective to weigh the effects of other intervening variables when system designers and managers attempt to predict response behaviors or implement training programs.

Although this study did not present strong evidence to suggest trust as the prime-mediating variable between system characteristics and response behavior, several limitations should be addressed. To accommodate for precise manipulations of the paradigm elements, this study was conducted under highly contrived laboratory conditions. It may be that trust toward the signaling system did not have adequate time to develop to a homeostatic state. Operators that interact with signaling systems generally do so over longer periods of time. Moreover, real world systems may not issue recommendations as frequently as the system used in this study. Therefore, it may be unrealistic to assume that trust developed in similar ways as those of the real world where it may be based on a few advisories over the course of hours, days, or weeks.

In conclusion, trust was not generally shown to be a strong causal variable in determining dependence behaviors, however, its role should not be dismissed. The more useful progression from studies that do not find trust to mediate signaling system relationships should be to empirically determine how trust interacts with other variables and place less predictive power within automation trust as a singular explanation. Empirical documentation of how these variables interact with trust would likely lead to the reconciliation of interpretational inconsistencies observed in the literature.

REFERENCES

Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior.* Upper Saddle River, NJ: Prentice Hall.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51* (6), 1173-1182.

Billings, C. E. (1996). *Human-centered aviation automation: Principles and Guidelines.* National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA.

Bliss, J. P. (2003). An investigation of extreme alarm response patterns in laboratory experiments. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* (pp. 1683-1687). Denver, CO.

Bliss, J. P., & Chancey, E. (2010). The effects of alarm system reliability and reaction training strategy on alarm responses. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting,* (pp. 2248-2252). San Francisco, CA.

Bliss, J. P., & Chancey, E. T. (under review) An investigation of training strategies to improve alarm reactions.

Bliss, J. P., & Dunn, M. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics, 43* (9), 1283-1300.

Bliss, J. P., Dunn, M., & Fuller, B. S. (1995). Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates. *Perceptual and Motor Skills, 38* (11), 1231-1242.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics*, *38* (11), 2300-2312.

Bliss, J. P., & Gilson, R. D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics*, *41* (1), 57-72.

Breznitz, S. (1984). *Cry Wolf: The Psychology of False Alarms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bustamante, E. A. (2008). Implementing likelihood alarm technology in integrated aviation displays for enhancing decision-making: A two-stage signal detection modeling approach. *International Journal of Applied Aviation Studies, 8* (2), 241-262.

Bustamante, E. A. (2009). A reexamination of the mediating effect of trust among alarm systems' characteristics and human compliance and reliance. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 249-253). Las Vegas, NV.

Carvalho, P. V., dos Santos, I. L., Gomes, J. O., Borges, M. R., & Guerlain, S. (2008). Human factors approach for evaluation and redesign of human-system interfaces of a nuclear power plant simulator. *Displays*, 273-284.

Chancey, E.T. & Bliss, J.P. (2012). Reliability of a Cued Combat Identification Aid on Soldier Performance and Trust. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting,* (pp. 1466-1470). Boston, MA.

Chancey, E.T. & Bliss, J.P. (2012). Unreliable information in infantry situation awareness: Improvement through game-based training. *Simulation and Gaming, 43* (5), 581-599.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*, *48* (3), 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, *49* (4), 564-572.

Drury, C. G., & Sinclair, M. A. (1983). Human and machine performance in an inspection task. *Human Factors*, *25* (4), 391-399.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, *13* (3), 147-164.

Elkin, P. L., Liebow, M., Bauer, B. A., Chaliki, S., Wahner-Roedler, D., Bundrick, J. et al. (2010). The introduction of a diagnostic decision support system (DXplain) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). *International Journal of Medical Informatics*, *79*, 772-777.

Endsley, M. R., & Kaber, D. K. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42* (3), 462-492.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analysis using

   G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research*

   Methods, 41 (4), 1149-1160.

Gaines, S. O., Panter, A. T., Lyde, M. D., Steers, W. N., Rusbult, C. E., Cox, C. L., et al.

   (1997). Evaluating the circumplexity of interpersonal traits and the manifestation

   of interpersonal traits in interpersonal trust. *Journal of Personality and Social*

   Psychology, 73, 610–623.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator

   response to warnings of danger: A laboratory investigation of the effects of the

   predictive value of a warning on human response time. *Journal of Experimental*

   Psychology: Applied, 1 (1), 19-33.

Gupta, N., Bisantz, A. M., & Singh, T. (2002). The effects of adverse condition warning

   system characteristics on driver performance: An investigation of alarm signal

   type and threshold level. *Behavioural and Information Technology, 21* (4), 235-

   248.

Halpin, S. M., Johnson, E. M., & Thornberry, J. A. (1973). Cognitive reliability in

   manned systems. *IEEE Transactions on Reliability, R-22* (3), 165-170.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new

   millennium. *Communication Monographs, 76,* 408-420.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable

   mediation, moderation, and conditional process modeling [White paper].

   Retrieved from http://www.afhayes.com/public/process2012.pdf

Hoaglin, D.C., & Iglewicz, B. (1987). Fine Tuning some resistant rules for outlier

   labeling. *Journal of American Statistical Association, 82,* 1147-1149.

Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically

   determined scale of trust in automated systems. *International Journal of*

   Cognitive Ergonomics, 4 (53), 53-71.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American*

   Psychologist, 39 (4), 341-350.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an

   "aid" can (and should) go unused. *Human Factors, 35* (2), 221-242.

Krupinski, E. A., Nodine, C. F., & Kundel, H. L. (1993). Perceptual enhancement of

   tumor targets in chest x-ray images. *Perception and Psychophysics, 53* (5), 519-

   526.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in

   human-machine systems. *Ergonomics, 35* (10), 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to

   automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance.

   Human Factors, 46 (1), 50-80.

Lees, M. N., & Lee, J. D. (2007). The influence of distraction and driving context on

   driver response to imperfect collision warning systems. *Ergonomics, 50* (8), 1264-

   1286.

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication-

   based explanation. *Communication Research, 31*(5), 499-523.

Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a

simulated navigation task. *Ergonomics, 50* (8), 1351-1364.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the

mediation, confounding, and suppression effect. *Prevention Science, 1*, 173-181.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A user's Guide* (2nd

Edition ed.). New York, NY: Lawrence Erlbaum Associates.

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* (4), 277-301.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2008). *Human

Performance: Cognition, stress and individual differences.* New York, NY:

Psychology Press.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A

Model Comparison Perspective* (2nd Edition ed.). New York, NY: Psychology

Press.

Meredith, C., & Edworthy, J. (1995). Are there too many alarms in the intensive care

unit? An overview of the problems. *Journal of Advanced Nursing, 21* (1), 15-20.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings.

Human Factors, 43 (4), 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human

Factors, 46* (2), 196-204.

Miller, C. A. (2002). Definitions and dimensions of etiquette. *Etiquette for human-computer work (Tech. Report FS-02-02, pp. 1-7)* . Menlo Park, CA: American Associates for Artificial Intelligence.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26* (4), 323-339.

Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors affecting trust in market research relationships. *Journal of Marketing, 57,* 81-101.

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* (1), 44-58.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27,* 527-539.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7* (3), 173-181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56* (1), 81-103.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology , 29* (5), 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27* (10), 864-876.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer

personalities be human personalities? *International Journal of Human-Computer

Studies, 43*, 223-239.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors

in Computing Systems*, (pp. 72-78). Boston, MA.

Norman, D. A., Ortony, A., & Russell, D. M. (2003). Affect and machine design:

Lessons for the development of autonomous machines. *IBM Systems Journal, 42*

(1), 38-44.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse,

abuse. *Human Factors, 39* (2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels

of human interaction with automation. *IEE Transactions on Systems, Man, and

Cybernetics - Part A: Systems and Humans, 30* (3), 286-297.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of

automation. *Human Factors, 50* (3), 511-520.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect

effects in simple mediation models. *Behavior Research Methods, Instruments, &

Computers, 36* (4), 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for

assessing and comparing indirect effects in multiple mediator models. *Behavior

Research Methods, 40* (3), 879-891.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93-115.

Pritchett, A. R. (2001). Reviewing the role of cockpit alerting systems. *Human Factors and Aerospace Safety, 1* (1), 5-38.

Rice, S. (2009). Examining single- and multiple-process theories of trust in automation. *The Journal of General Psychology, 13* (3), 303-319.

Riley, V. (1994). A theory of operator reliance on automation. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Recent Research and Trends* (pp. 8-14). Hillsdale, NJ: Erlbaum.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*, 651–665.

Safar, J. A., & Turner, C. W. (2005). Validation of a two factor structure for system trust. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, (pp. 497-501). Orlando, FL.

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

Sarter, N. B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43*(4), 573-583.

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *The International Journal of Aviation Psychology, 2* (4), 303-321.

Sarter, N. B., & Woods, D. D. (2000). Team play with a powerful and independent agent: A full-mission simulation study. *Human Factors, 42* (3), 390-402.

Selwyn, N., Gorard, S., Furlong, J., & Madden, L. (2003). Older adults' use of information and communications technology in everyday life. *Ageing and Society, 23*(5), 561-582.

Sheridan, T. B., & Hennessy, R. T. (1984). *Research and modeling of supervisory control behavior.* Washington, DC: National Academy.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype Susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422-445.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84* (3), 1107-1108.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30* (4), 445-459.

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

Steele, J. R., & Ambady, N. (2006). "Math is Hard!" The effect of gender priming on

women's attitudes. *Journal of Experimental Social Psychology, 42*, 428-436.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated

combat identification system. *Human Factors, 51* (3), 281-291.

Webb, C. M., & Hewett, K. J. (2010). *An analysis of U.S. Army fratricide incidents

during the Global War on Terror (11 September 2001 to 31 March 2008).* United

States Army Aeromedical Research Laboratory - Warfighter Performance and

Health Division.

Wickens, C. D. (1998). Commonsense statistics. *Ergonomics in Design, 6* (4), 18-22.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation:

A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8* (3),

201-212.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human

Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Lee, J., Liu, Y., & Becker, S. G. (2004). *An Introduction to Human

Factors Engineering* (2nd ed.). Upper Saddle River, NY: Prentice Hall.

Wiczorek, R., & Manzey, D. (2010). Is operators' compliance with alarm systems a

product of rational consideration? *Proceedings of the Human Factors and

Ergonomics Society 54th Annual Meeting,* (pp. 1722-1726). San Francisco, CA.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects

of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomic

Science, 2* (4), 352-367.

APPENDIX A

INFORMED CONSENT FORM

INFORMED CONSENT DOCUMENT

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.

TITLE OF RESEARCH: User Performance with an Alarm Based Task

RESEARCHERS:

James P. Bliss, Ph.D., Professor, Responsible Project Investigator, College of Sciences, Psychology Department
Eric T. Chancey, graduate student, College of Sciences, Psychology Department.

DESCRIPTION OF RESEARCH STUDY:

Auditory and visual alarm signals are often used in complex task environments to warn operators of dangerous or abnormal conditions. However, operators often use inefficient strategies when presented with alarm systems that behave unreliably. Research has suggested various reasons for these behaviors however more investigation is needed.

Sixty participants will be tested in this experiment. Those who agree to be tested will complete a background information form. Following this, you will be asked to perform a familiarization session with multiple tasks that simulate tasks similar to those that aircraft pilots perform. After training, you will be asked to perform the simulated aircraft tasks while also interacting with an alarm system in three experimental sessions. During the sessions, you will monitor a computer screen that depicts gauges that need to be monitored. When an alarm sounds, you will decide how to react to it (respond or ignore) using a computer keyboard. You will also be asked questions regarding your attitude toward the system. After three experimental sessions, you will complete an opinion questionnaire to indicate your strategy for responding. You will then be debriefed and dismissed. The entire experiment should last approximately 2 hours.

EXCLUSIONARY CRITERIA:

To participate, you must be over the age of 18. You must have normal vision or corrected-to-normal vision. You must also have normal or corrected-to-normal hearing. Therefore, if you normally wear eyeglasses, contact lenses or hearing aids you will need to wear them to participate.

RISKS AND BENEFITS:

RISKS: If you decide to participate in this study, then you may face a risk of eyestrain similar to the eyestrain experienced during normal computer usage. The researcher tried to reduce these risks by limiting the experimental participation time to less than two hours. 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 direct benefits for participation in this study. However, you may learn valuable information about how research is conducted.

COSTS AND PAYMENTS:

The researchers want your decision about participating in this study to be absolutely voluntary. The main benefit to you for participating in this study is the extra credit or course credit points that you will earn for your class. Although they are unable to give you payment for participating in this study, if you decide to participate in this study, you will receive 2 Psychology Department research credits, which may be applied to course requirements or extra credit in certain Psychology courses. Equivalent credits may be obtained in other ways. You do not have to participate in this study, or any Psychology Department study, to obtain this credit.

CONFIDENTIALITY:
Your participation is completely confidential. The researcher will remove all identifiers from the information. The results of this study may be used in reports, presentations, and publications; but the researcher will not identify you individually in such publications.

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 neither affect your relationship with Old Dominion University, nor 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 agree to participate, then your consent in this document does not waive any of your legal rights. However, in the event of harm, injury, or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact Dr. James P. Bliss at 757-683-4051, Dr. George Maihafer (IRB Chair) at 757-683-4520, or the ODU Office of Research, 757-683-3460.

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, please contact the researcher at the number above.

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. George Maihafer (IRB Chair) from the Old Dominion University Office of Research, 757-683-4520, or the ODU Office of Research, 757-683-3460.

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.

---------------------------------   -------------------------------------   ------------------------

Participant's Name   Participant's Signature   Date

---------------------------------   -------------------------------------   ------------------------

Investigator's Name   Investigator's Signature   Date

APPENDIX B

DEMOGRAPHIC FORM

PARTICIPANT BACKGROUND INFORMATION FORM   Participant #_____

Date:_____ Time:_____

The purpose of this questionnaire is to collect background information for participants in this experiment. This information will be used strictly for this experiment and for research purposes only. Please complete or circle each item to the best of your knowledge.

1. Age _____

2.   Male
      Female

3. Have you ever been diagnosed as color blind or color deficient? _____

      0 = No
      1 = Yes
4. Have you ever been diagnosed as having hearing loss?_____

      0 = No
      1 = Yes
5. If yes, do you have correction with you (i.e. hearing aid)?_____

      0=No
      1=Yes
6. Have you ever been diagnosed as being nearsighted (myopic)? _____

      0=No
      1=Yes

7. Have you ever been diagnosed as being farsighted (hyperopic)?_____

      0=No
      1=Yes

8. If you answered yes to either #6 or #7, do you have correction with you (i.e. glasses, contact lenses, etc.)? _____

      0=No
      1=Yes

9. How many hours per week do you play video/simulation games? _____

10. How many hours per week do you use a computer (work and recreation combined)?_____

APPENDIX C

INSTRUCTION SHEET

Welcome to project CALIBRATION, the following experiment should take 2 hours, at the end of which you will receive 2 SONA credits. Please put away your cell phone and turn it on silent. You will be asked to take part in several tasks using a computer program called the MATB. The tasks in MATB simulate the kinds of tasks that pilots perform during flight. The tasks you will be expected to respond to will be the Resource management and tracking tasks. These tasks must be completed the entire time and should be considered the primary task.

Tracking Task

The upper central region of the MATB window contains the tracking task. Your job is to keep the target in the center of the rectangular box. The overall purpose of this task is to keep the aircraft (represented by the blue circle) within the dotted rectangular area in the center of this task. Try to maintain this at all times. You control the aircraft with movements of the joystick. If you do not control the aircraft with the joystick, it will drift away from the center. If the aircraft leaves the rectangular area try to bring the aircraft back to center as quickly as possible.
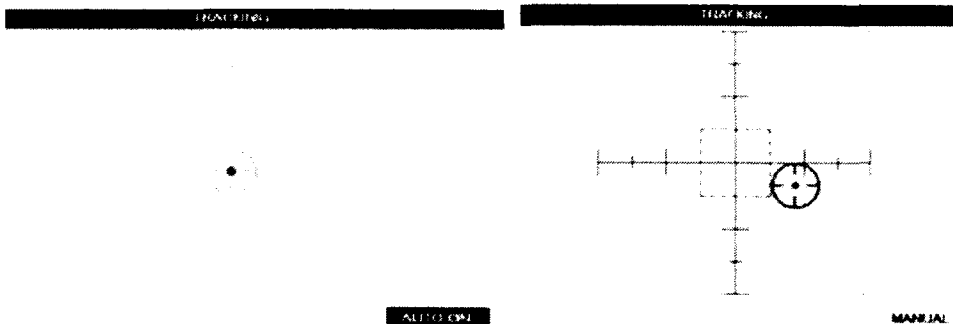
Resource Management task

The lower right region of the MATB-II main window contains the resource management task. The rectangular regions identified with the letters A-F represent fuel tanks. The green levels within the tanks represent fuel levels. Along the lines which connect the tanks are pumps which transfer fuel from one tank to another in the direction indicated by the arrows.

There are 8 pumps labeled with the numbers 1-8. Each one of the pumps is represented by a rectangular box with a number inside it that identifies the pump, and an arrow that indicates the direction of the fuel. The pumps are used to transfer fuel from the supply tanks to the main tanks.

Deactivated pumps are colored in gray , activated pumps are green , and failed pumps are red ▐�. Note in the figure that pumps 1, 2, 4, and 6 are active, pumps 3, 7, and 8 are inactive, and pump 5 is failed.
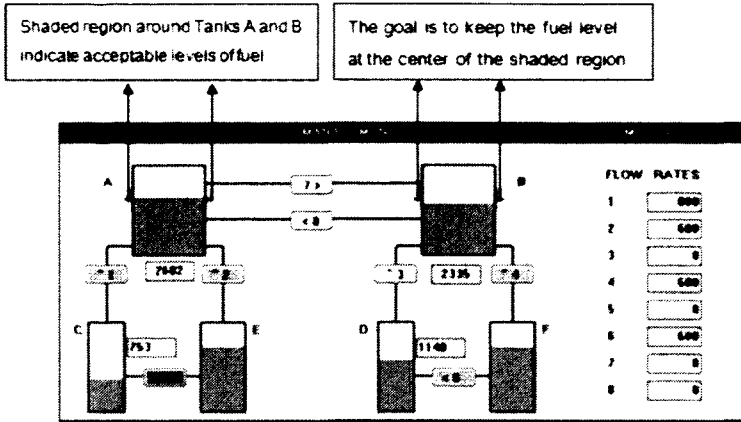
When a pump activates, the numbers change in the "Pump Status" area. Under "Pump Status," two columns of numbers are present. In the first column, numbers 1 through 8, correspond directly to the pumps in the diagram. The second column indicates the flow rate in units per minute for each pump when it is on.

In the figure below, the numbers underneath tanks A and B and to the right on tanks C and D represent the amount of fuel for each of those tanks. Those numbers will be increasing and decreasing as the fuel levels change. The capacity for the main tanks, A and B, is 4000 units each. The supply tanks, C and D, contain a maximum of 2000 units each. Tanks E and F are supply tanks that have an unlimited capacity – they never run out. The areas shaded in light blue on the side of tanks A and B indicate the critical levels of fuel for those tanks. You must transfer fuel to tanks A and B to meet these criteria because the fuel tanks A and B are always being consumed.
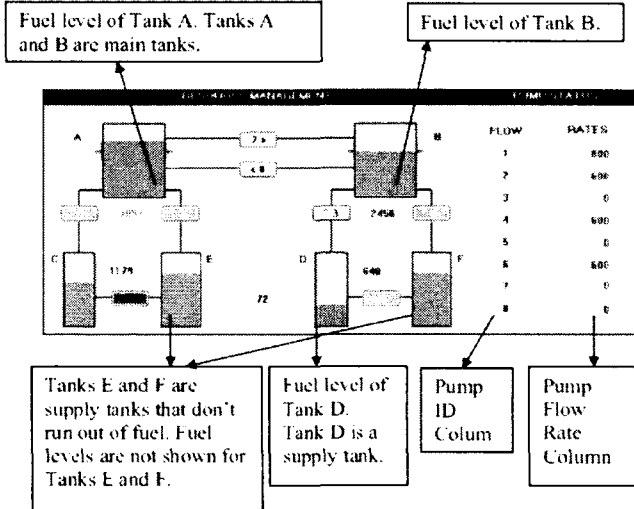
When the resource management task begins, the fuel level for Tanks A and B is at 2500 units. You are to keep the level of fuel from dropping below this level as indicated by the marker on either side of these pumps. As time passes, tanks A and B lose fuel. These tanks would eventually become empty without the transfer of additional fuel. Tanks C and D only lose fuel if they are transferring fuel to another tank.

Let's consider the process of transferring fuel. Each pump can only transfer fuel in the direction indicated by the ^ arrow in its label. The pumps are activated by pressing the number key corresponding to the pump that you wish to activate. A pump is actively transferring fuel when it turns green.

So far, you've seen two conditions for the pumps: ON and OFF. If you press the pump number on the keyboard just once, you will turn the pump ON ; pressing the key again turns that pump OFF

[⌐ ²], and so on. If a tank fills up to its capacity, all incoming pump lines will be turned off automatically. This is because a full tank cannot receive any more fuel. You will have to turn those pumps back on at a later time, if the fuel level of the tank goes below the critical level. Furthermore, if a tank becomes empty, all outgoing pumps will automatically be turned off. This is because an empty tank can no longer transfer fuel. In that case, the proper action is to supply fuel to an empty tank before turning a pump that transfers fuel out of it.

At some point during the execution of the resource management task, one or more of the pumps may fail. When a pump fails, its label turns red. Depending on the level of fuel in the tank affected, you might need to transfer fuel from one main tank to another main tank to compensate for the loss of fuel. You can cross feed fuel from one main tank to the other by activating either pump 7 or 8 (see the figure below)
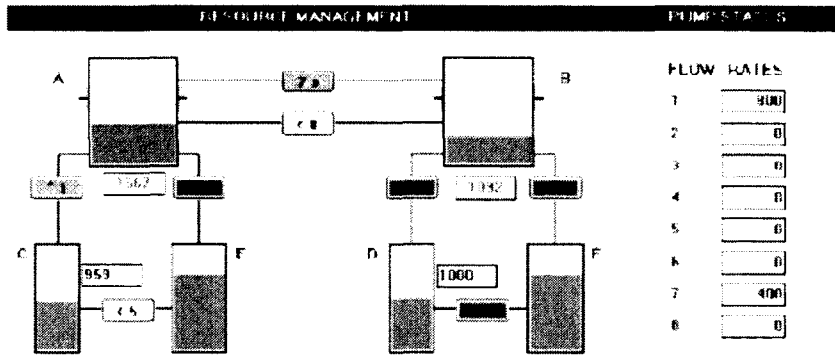
Once again, the overall goal is to maintain the fuel level in tanks A and B close to 2500 units each for as long as you possibly can. There may be more than one way to achieve this goal; you may use the method that works best for you. If the fuel level in these tanks should deviate from this level, please return the fuel level back this this point as soon as possible.

Signaling System Task

In addition to the MATB tasks, you will be asked to interact with an alarm system that will signal changes in 5 different gauges. Although this is a secondary task, these gauges must remain within a critical threshold. The alarm system will indicate a problem if one of these gauges exceeds this critical threshold. However, it is important to realize that this system fluctuates in its reliability or accuracy in indicating a breach of this critical threshold. Sometimes this alarm system is relatively reliable and will very accurately indicate that the gauges are within normal operating conditions. However, sometimes the probability of this system accurately indicating that a gauge has breached the critical threshold will be a little better than chance. You will be asked to interact with this system in the most efficient strategy you deem possible. To help you with this strategy I will inform you of the reliability of the system prior to your interaction with the system in all experimental session. The system will not change in its reliability during each experimental session. Do you have any questions?

If the alarm system thinks that one of the gauges has breached the critical threshold then it will indicate a system failure. This is illustrated in the image below.
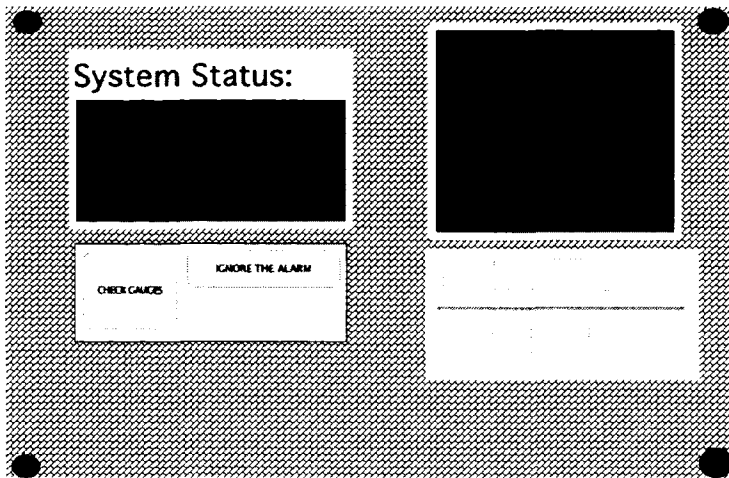
If presented with this system status you have two choices, you may: click on the **CHECK GAUGES** button and fix the potential problem OR you may click on the **IGNORE THE ALARM** button and continue interacting with the MATB tasks.

If the alarm system thinks that all of the gauges are within normal conditions (and not exceeding the critical threshold) then it will indicate that the system is OK. This is illustrated in the image below.
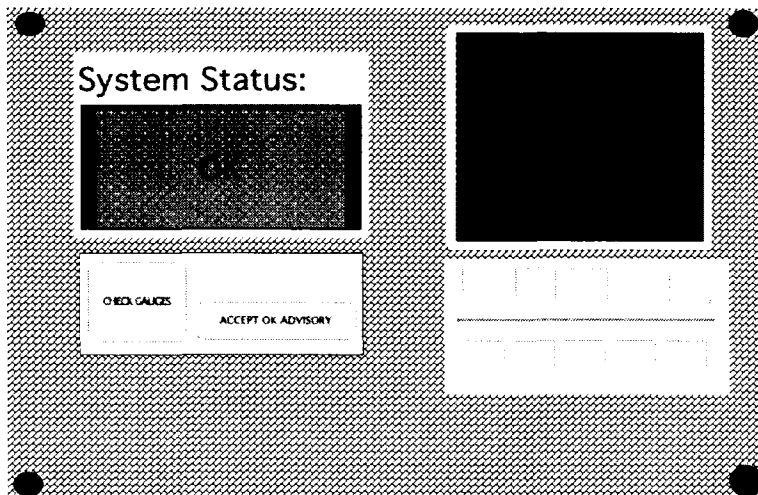
If presented with this system status you have two choices, you may: click on the **CHECK GAUGES** button and investigate if there is a potential problem OR you may click on the **ACCEPT OK ADVISORY** button and continue interacting with the MATB tasks.

If you do decide to click on the **CHECK GAUGES** button you will be presented with the 5 different gauges.

If the system is in normal operating conditions then all of the yellow bars will be between the green lines. If this is the case then you must click the CHANGE VIEW button located in the bottom right corner of the screen to move on.

If the system is not within normal operating conditions then one of the bars will be above or below the green lines. If this is the case then you must click on the button below the gauge that has breached the critical threshold to move on.

What should you do when presented with both of these illustrations below?
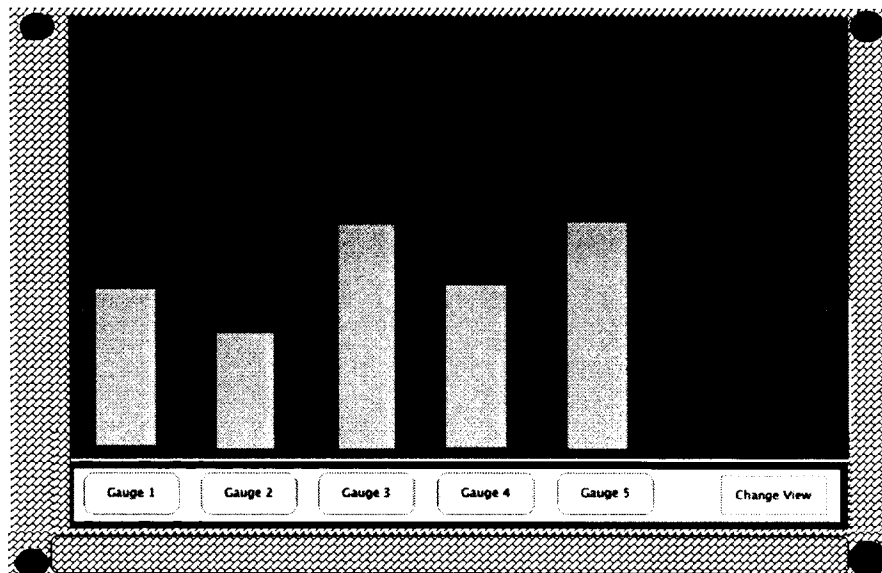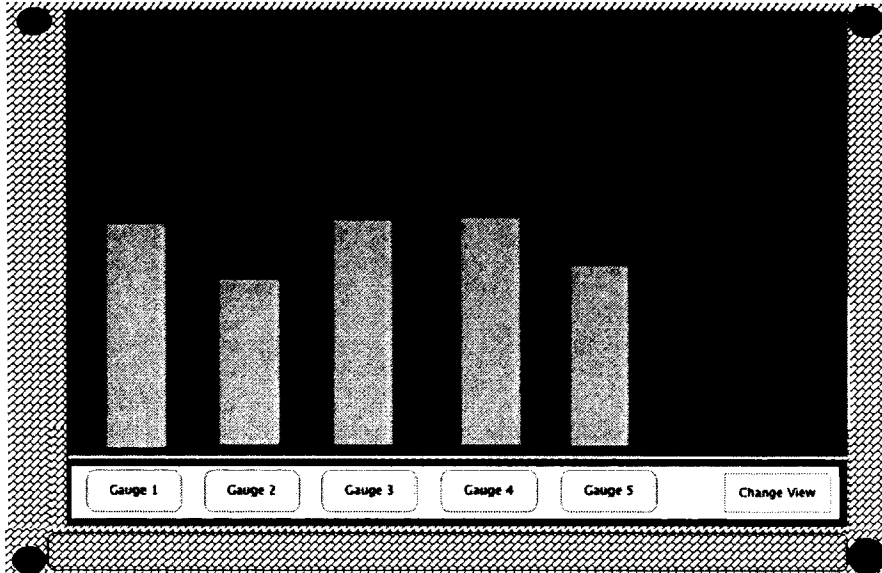
When interacting with this system it is important that you only click the **CHECK GAUGES** button if you suspect that there is a gauge that needs to be fixed. If you *appropriately* click this button, indicating that there is an actual gauge needing to be fixed, you will hear a voice that says "Correct" and receive 1 point to a bank which will be displayed at the bottom of the screen. If you *inappropriately* click this button, when no gauges need to be fixed, you will hear a voice that says "Incorrect" and 1 point will be deducted from your bank.

It is also important that you only click the **ACCEPT OK ADVISORY** and **IGNORE THE ALARM** buttons if you suspect that none of the gauges needs to be fixed. If you *appropriately* click these buttons, indicating that no gauges actually needed to be fixed, then you will hear a voice that says "Correct" and then receive 1 point toward your bank. If you *inappropriately* click one of these buttons, when a gauge actually did need to be fixed, then you will hear a voice that says "Incorrect" and 1 point will then be deducted from your bank. Do you have any questions about this?

For researcher only:

For this experimental session, we know from past performance history that the alarm which is on the computer to your left is **(60%, 75%, 90%)** reliable, in other words – out of every 10 advisory signals **(6; between 7 and 8; 9)** of those 10 advisory signals will be correct. Therefore, the remaining **(4; 2 or 3; 1)** advisory signals will be incorrect. So again, the system is about **(60%, 75%, 90%)** reliable. You should also know that when this system makes a mistake it tends to **(miss; falsely indicate)** when one of the gauges needs to be checked. You should use this information to help you decide which alarms to respond to, and which to cancel. Do you have any questions about the reliability of this system or any other general questions about the experiment so far? Ok just to make sure that you understand the characteristics of this alarm system I am going to ask you a fill out this brief questionnaire.

APPENDIX D

TRUST QUESTIONNAIRE FORM

Part. #:_____ Group:_____ Session:_____ Date:_____ Time:_____

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

1= not descriptive statement of targeting aid impression
12= very descriptive statement of targeting aid impression

1.) The alarm system is deceptive

   1   2   3   4   5   6   7   8   9   10   11   12

2.) The alarm system behaves in an underhanded manner

   1   2   3   4   5   6   7   8   9   10   11   12

3.) I am suspicious of the alarm system's outputs

   1   2   3   4   5   6   7   8   9   10   11   12

4.) I am wary of the alarm system

   1   2   3   4   5   6   7   8   9   10   11   12

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

   1   2   3   4   5   6   7   8   9   10   11   12

6.) I am confident in the alarm system

   1   2   3   4   5   6   7   8   9   10   11   12

7.) The alarm system provides security

   1   2   3   4   5   6   7   8   9   10   11   12

8.) The alarm system has integrity

   1   2   3   4   5   6   7   8   9   10   11   12

9.) The alarm system is dependable

   1   2   3   4   5   6   7   8   9   10   11   12

10.) The alarm system is reliable

   1   2   3   4   5   6   7   8   9   10   11   12

11.) I can trust the alarm system will accurately indicate problems to respond to

   1   2   3   4   5   6   7   8   9   10   11   12

APPENDIX E

OPINION QUESTIONNAIRE FORM

Participant No._____ Date:_____ Time:_____

Please answer the following questions about yourself by circling the most appropriate response. The information you provide will be kept completely confidential and will not be linked backed to you in any way.

Please circle only one answer per question.

1. This experiment was time consuming.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

2. This experiment was confusing.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

3. I did not feel like I had a good grasp on the instructions for this experiment.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

4. I feel like I performed well on this experiment.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

5. I feel like I performed poorly on this experiment.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

6. This experiment was easy to understand

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

7. This experiment was enjoyable.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

8. I did not enjoy this experiment.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

9. I am glad that I participated in this experiment

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

10. I felt engaged in the tasks for this experiment.

      Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

11. I felt like I received adequate time to train and get comfortable with the experimental task before

beginning the actual experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

12. I felt like I did not receive adequate time to train and get comfortable with the experimental task before beginning the actual experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

13. I felt motivated to perform to the best of my ability in this experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

14. I did not care how well I performed in this experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

15. I tried my best to perform well on this experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

16. I did not try my best to perform well on this experiment.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly

17. Overall, I would recommend this experiment to other students.

Disagree strongly   Disagree   Neutral   Agree   Agree Strongly
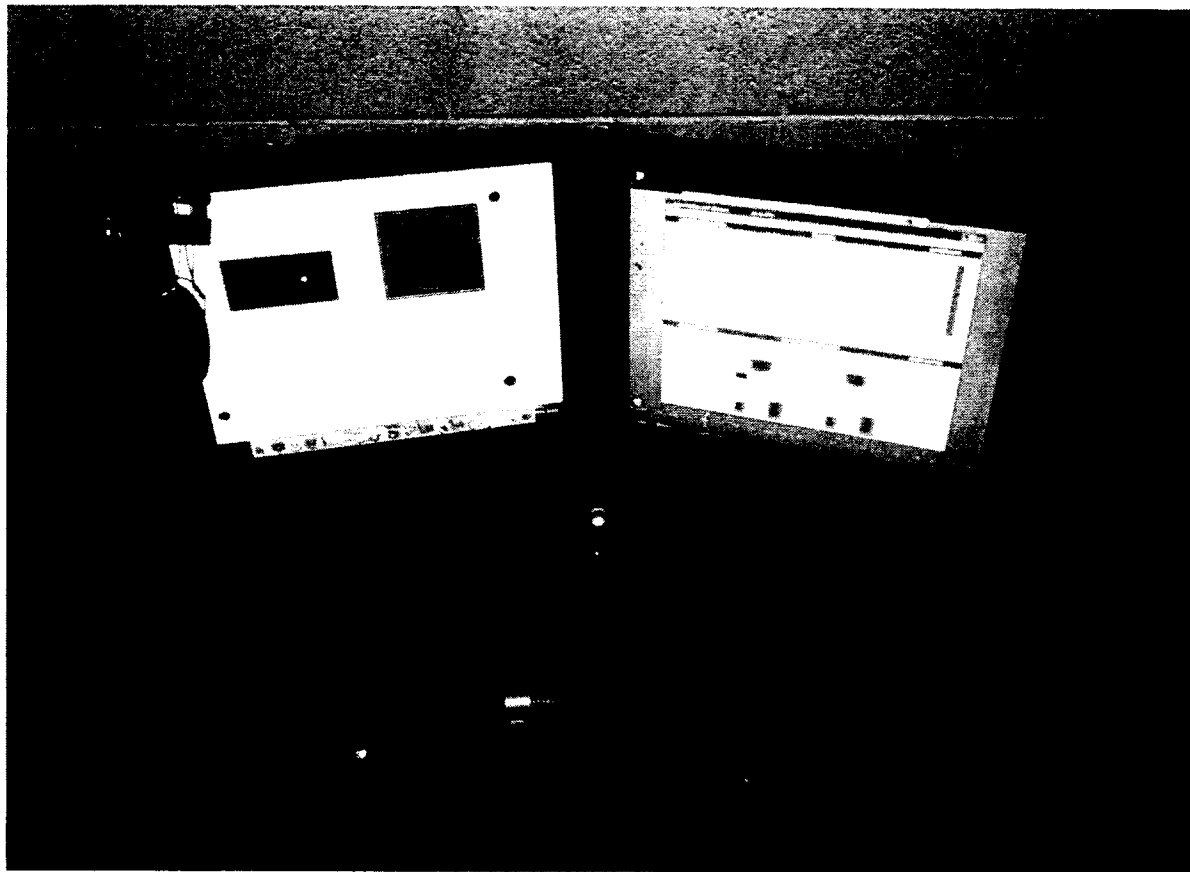
18. Did you have a strategy for responding to the experimental task?

Yes   No

If yes, please describe

_____

_____

_____

19. Do you have any other thought, feelings, or comments about the experiment?

_____

_____

_____

APPENDIX F

PICTURE OF EXPERIMENTAL SETUP

APPENDIX G

SECONDARY SIGNALING SYSTEM TASK BREAKDOWN

	100%	90% MP	60% MP	90% FP	60% FP
1	H	H	H	H	FA
2	H	M	CR	CR	FA
3	CR	H	CR	H	CR
4	CR	CR	M	CR	FA
5	CR	CR	CR	H	H
6	H	CR	M	CR	H
7	H	CR	M	FA	H
8	CR	H	CR	H	H
9	CR	CR	M	H	FA
10	H	H	CR	CR	H
11	CR	H	H	FA	H
12	H	H	CR	H	H
13	H	CR	M	H	H
14	H	CR	M	H	H
15	CR	CR	CR	CR	FA
16	H	M	CR	CR	CR
17	CR	CR	M	CR	FA
18	H	CR	M	CR	FA
19	CR	H	CR	H	H
20	CR	H	CR	H	FA
21	H	CR	CR	CR	FA
22	CR	H	M	CR	H
23	CR	CR	M	H	H
24	H	CR	H	CR	FA
25	H	CR	CR	H	H
26	H	CR	M	H	FA
27	CR	H	M	H	CR
28	CR	H	CR	H	H
29	CR	M	CR	FA	FA
30	H	H	CR	CR	H
		TRUS QUEST.			
31	H	H	H	H	FA
32	CR	H	CR	H	FA
33	H	CR	CR	CR	H
34	H	CR	M	H	FA
35	CR	CR	CR	CR	H
36	H	H	M	CR	FA
37	H	CR	M	FA	H
38	CR	M	CR	H	H
39	CR	H	M	CR	CR
40	CR	CR	CR	H	H
41	H	H	M	CR	FA
42	CR	CR	CR	H	H
43	H	CR	M	FA	H
44	CR	H	CR	CR	FA
45	CR	CR	CR	CR	CR
46	H	M	M	H	H
47	H	CR	CR	H	H
48	CR	CR	CR	H	FA
49	CR	H	M	CR	H
50	H	H	H	H	FA
51	H	CR	CR	CR	FA
52	H	M	M	CR	H
53	CR	CR	CR	H	H
54	CR	H	M	H	H
55	CR	CR	H	H	H
56	H	CR	CR	H	FA
57	CR	H	M	FA	CR
58	CR	H	CR	H	H
59	H	CR	M	CR	FA
60	H	H	CR	CR	FA

H= hit / signaling system correctly indicates a system failure

FA = false alarm / signaling system incorrectly indicates a system failure

M = miss / signaling system incorrectly indicates system OK

CR = correct rejection / signaling system correctly indicates system OK

After 30th signal = administration of trust questionnaire

31st-60th signal = response data to these signals will be analyzed

100% BL – 30H, 0 FA, 0 M, 30 CR

90% FP – 30 H, 6 FA, 0 M, 24 CR
60% FP – 30 H, 24 FA, 0 M, 6 CR

90% MP – 24 H, 0 FA, 6 M, 30 CR
60% MP – 6 H, 0 FA, 24 M, 30 CR

Or proportionally represented as

100% BL - .50 H, .00 FA, .00 M, .50 CR

90% FP – .50 H, .10 FA, .00 M, .40 CR
60% FP – .50 H, .40 FA, .00 M, .10 CR

90% MP – .40 H, .00 FA, .10 M, .50 CR
60% MP – .10 H, .00 FA, .60 M, .50 CR

50% signal and 50% noise

APPENDIX C

SONA SIGN UP SHEET

User Performance with an Alarm Based Task

James P. Bliss and Eric T. Chancey, of the ODU Psychology Department are currently conducting an experiment.

<u>Brief research overview:</u> The purpose of this research is to investigate how operators react to alarm systems with different characteristics.

<u>Research overview</u>

      Auditory and visual alarm signals are often used in complex task environments to warn operators of dangerous or abnormal conditions. However, operators often use inefficient strategies when presented with alarm systems that behave unreliably. Research has suggested various reasons for these behaviors however more investigation is needed.

      Sixty participants will be tested in this experiment. Those who agree to be tested will complete a background information form. Following this, you will be asked to perform a familiarization session with multiple tasks that simulate tasks similar to those that aircraft pilots perform. After training, you will be asked to perform the simulated aircraft tasks while also interacting with an alarm system in four experimental sessions. During the sessions, you will monitor a computer screen that depicts gauges that need to be monitored. When an alarm sounds, you will decide how to react to it (respond or ignore) using a computer keyboard. After four experimental sessions, you will complete an opinion questionnaire to indicate your strategy for responding. You will then be debriefed and dismissed. The entire experiment should last approximately 2 hours.

You will receive a $5 Starbucks *or* Chick-fil-A gift card and 2 Psychology Department research credit.

APPENDIX I

ERROR CHARACTERISTICS MANIPULATION CHECK FORM

Participant_____ Session_____ G_____ R_____

1.) If the alarm system issued 100 system advisory signals, how many times would it correctly

indicate that there is either a gauge that needs to be fixed (yellow bar is outside of green lines) or

that all of the gauges are in a normal state (yellow bar is between green lines)? Please draw a line

to indicate your answer.

```
+--------+--------+--------+--------+--------+--------+--------+--------+--------+-------+
0       10       20       30       40       50       60       70       80              90      100
```

2.) How is this alarm system likely to behave?

    a.   Fail to indicate when there is a gauge that needs to be fixed

    b.   Incorrectly indicate that there is a gauge that needs to be fixed

3.) How many times is this system likely to make an error during the experimental session?

    a.   1 out of every 4 signals issued

    b.   10 out of every 100 signals issued

    c.   4 out of every 10 signals issued

    d.   2 out of every 10 signals issued

APPENDIX J

DISPOSITIONAL TRUST QUESTIONNAIRE

Part. #: _____ Group: _____ Date: _____ Time: _____

Below is a list of statements for evaluating trust between people and systems.
Please circle the number that best describes how you feel toward each of these systems

1= not descriptive statement
12= very descriptive statement

The smoke alarm where I live:

1.) I am confident in the smoke alarm

    1   2   3   4   5   6   7   8   9   10   11   12

2.) The smoke alarm is dependable

    1   2   3   4   5   6   7   8   9   10   11   12

3.) The smoke alarm is reliable

    1   2   3   4   5   6   7   8   9   10   11   12

4.) I can trust the smoke alarm will accurately indicate if there is a fire or smoke that I need to respond to

    1   2   3   4   5   6   7   8   9   10   11   12

The check engine gauge in most cars:

5.) I am confident in the check engine gauge

    1   2   3   4   5   6   7   8   9   10   11   12

6.) The check engine gauge is dependable

    1   2   3   4   5   6   7   8   9   10   11   12

7.) The check engine gauge is reliable

    1   2   3   4   5   6   7   8   9   10   11   12

8.) I can trust the check engine gauge will accurately indicate problems in the engine that should be responded to

    1   2   3   4   5   6   7   8   9   10   11   12

My alarm clock:

9.) I am confident in my alarm clock

1   2   3   4   5   6   7   8   9   10   11   12

10.) My alarm clock is dependable

1   2   3   4   5   6   7   8   9   10   11   12

11.) My alarm clock is reliable

1   2   3   4   5   6   7   8   9   10   11   12

12.) I can trust my alarm clock will accurately go off when I have set it to wake me up

1   2   3   4   5   6   7   8   9   10   11   12

The computer I use most frequently

13.) I am confident in the computer I use most frequently

1   2   3   4   5   6   7   8   9   10   11   12

14.) The computer I use most frequently is dependable

1   2   3   4   5   6   7   8   9   10   11   12

15.) The computer I use most frequently is reliable

1   2   3   4   5   6   7   8   9   10   11   12

16.) I can trust that the computer I use most frequently will accurately give me the information I request

1   2   3   4   5   6   7   8   9   10   11   12

Modern technology in general (GPSs, computers, cell phones, vehicles, TVs, etc.)

17.) I am confident in modern technology in general

1   2   3   4   5   6   7   8   9   10   11   12

18.) Modern technology in general is dependable

1   2   3   4   5   6   7   8   9   10   11   12

19.) Modern technology in general is reliable

1   2   3   4   5   6   7   8   9   10   11   12

20.) I can trust that modern technology in general accurately does what it is designed to do

1   2   3   4   5   6   7   8   9   10   11   12

APPENDIX K

DESCRIPTIVE STATISTICS FOR EXPERIMENTAL DATA

<u>**Sensitivity**</u>

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
90%	44	2.30	2.12	2.49	0.09
60%	44	0.39	0.28	0.49	0.05
Error Bias					
False Alarm Prone	22	1.37	1.22	1.53	0.08
Miss Prone	22	1.32	1.16	1.48	0.08

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
Reliability					
Baseline (100%)	43	3.03	2.70	3.36	0.16
90%	43	2.34	2.14	2.52	0.09
60%	43	0.38	0.27	0.49	0.05

Response Bias

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
90%	44	-0.09	-0.18	-0.00	0.05
60%	44	-0.09	-0.23	0.44	0.07
Error Bias					
False Alarm Prone	22	-0.31	-0.43	-0.19	0.06
Miss Prone	22	0.13	0.01	0.25	0.06

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
Reliability					
Baseline (100%)	43	-0.03	-0.09	0.02	0.03
90%	43	-0.09	-0.21	0.04	0.06
60%	43	-0.10	-0.25	0.05	0.07

Agreement Rate

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
90%	44	55.71	54.11	57.30	0.79
60%	44	38.43	35.54	41.33	1.44
Error Bias					
False Alarm Prone	22	48.21	45.85	50.56	1.17
Miss Prone	22	45.93	43.57	48.29	1.17

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
Reliability					
Baseline (100%)	43	54.54	52.65	56.42	0.94
90%	43	55.81	54.16	57.47	0.82
60%	43	38.33	35.39	41.26	1.45

Reaction Time (secs)

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
90%	44	2.72	2.49	2.95	0.11
60%	44	2.86	2.63	3.09	0.11
Error Bias					
False Alarm Prone	22	2.81	2.63	3.09	0.11
Miss Prone	22	2.72	2.50	2.95	0.11

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
Reliability					
Baseline (100%)	43	3.51	3.13	3.88	0.19
90%	43	2.70	2.47	2.93	0.11
60%	43	2.83	2.60	3.05	0.11

Signaling System Trust

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals Min	Max	SE
90%	44	85.27	78.56	91.99	3.33
60%	44	40.25	34.94	45.57	2.63
Error Bias					
False Alarm Prone	22	66.32	60.47	72.17	2.99
Miss Prone	22	59.21	53.35	65.06	2.99

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals Min	Max	SE
Reliability					
Baseline (100%)	43	96.26	86.69	105.83	4.74
90%	43	85.65	78.76	92.55	3.42
60%	43	40.14	34.71	45.57	2.69

Signaling System Task Score

Data used for split-plot analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
90%	44	61.86	59.11	64.62	1.36
60%	44	28.05	25.96	30.13	1.03
Error Bias					
False Alarm Prone	22	46.23	43.80	48.66	1.21
Miss Prone	22	43.68	41.25	46.11	1.21

Note: Data are collapsed across each other

Data used for one-way repeated measure analyses

Source	n	Mean	95% Confidence Intervals		SE
			Min	Max	
Reliability					
Baseline (100%)	43	69.02	65.26	72.79	1.87
90%	43	62.14	59.30	64.98	1.41
60%	43	27.91	25.82	30.00	1.04

VITA

Eric T. Chancey

Department of Psychology   Tel: (540) 842-7254
250 Mills Godwin Building   Email: echan004@odu.edu
Old Dominion University
Norfolk, VA 23529 - 0267

EDUCATION

(2010-2015) Ph.D.   Old Dominion University, Norfolk, VA
                           Human Factors Psychology (In Progress)
2010 B.S.   Old Dominion University, Norfolk, VA
                           Major: Psychology
                           Minor: Business Management
                           Magna Cum Laude
                           Department Honors

PEER REVIEWED PUBLICATIONS

Chancey, E.T. & Bliss, J.P. (2012). Unreliable information in infantry situation awareness: Improvement through game-based training. *Simulation and Gaming, 43* (5), 581-599.

PEER REVIEWED PAPERS PRESENTED AT PROFESSIONAL MEETINGS

Chancey, E.T. & Bliss, J.P. (2012). Reliability of a cued combat identification aid on Soldier performance and trust. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, (pp. 1466-1470). Boston, MA.

Bliss, J. P., & Chancey, E. (2010). The effects of alarm system reliability and reaction training strategy on alarm responses. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, (pp. 2248-2252). San Francisco, CA.

NON-PEER REVIEWED PUBLICATIONS

Bliss, J.P., Liebman, R., Chancey, E.T. (2012). Training time estimation to improve alarm reactions. *Work, 41*, 3580-3585.

BOOK CHAPTERS

Bliss, J. P., Proaps, A., & Chancey, E. T. (under review). Human Performance Measurement in Virtual Environments. In K. M. Stanney, & K. S. Hale (Eds.), *Handbook of Virtual Environments: Design, Implementation, and Application* (2nd ed.).

PAPERS UNDER REVIEW

Bliss, J. P., & Chancey, E. T. (under review) An investigation of training strategies to improve alarm reactions.

STUDENT CONFERENCE PRESENTATIONS

Chancey, E.T. & Bliss, J.P. (2012). The Effects of Reliability of a Cued Combat Identification System on Shooter Performance. Research Presented at the Virginia HFES Student Chapter Conference, Norfolk, VA.

AWARDS

Undergraduate (Old Dominion University):
        Academic Achievement Scholarship, Old Dominion University - 2009-2010. Maintain an honorable cumulative GPA between 4.0 and 3.85. $2,000
        Dean's List –Spring 2009, Fall 2009, Spring 2010

AFFILIATIONS

Human Factors and Ergonomics Society ODU Student Chapter   October 2009
        Treasurer   May 2012 – Present
Human Factors and Ergonomics Society National Chapter   September 2010
Golden Key Honor Society   November 2009
Phi Kappa Phi   April 2010