Winter 2014

Effects of Simultaneous Alarms on Resolution Heuristics

Amanda C. Allen
Old Dominion University

EFFECTS OF SIMULTANEOUS ALARMS ON RESOLUTION HEURISTICS

by

Amanda C. Allen
B.A. May 2010, Clemson University

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

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2014

Approved by:

_____
J. Christopher Brill (Director)

_____
James P. Bliss (Member)

_____
Poornima Madhavan (Member)

ABSTRACT

EFFECTS OF SIMULATANEOUS ALARMS ON RESOLUTION HEURISTICS

Amanda C. Allen
Old Dominion University, 2014
Director: Dr. J. Christopher Brill, Ph.D.

Automated signaling systems are frequently used to direct operator attention to potential hazards. Although these automated systems can lead to enhanced human performance, factors such as degraded alarm signal reliability and lack of trust can undermine the potential benefits of automation (Breznitz, 1984; Rice, 2009, Wickens & 2007). Additionally, work by Gilson, Mouloua, Graft, and McDonald (2001), as well as Keller and Rice (2009), suggest that an alarm contained within a larger array of alarms should not be evaluated individually. Due to the increasing use of multiple alarms in complex environments such as operating rooms and cockpits (Konkani, Oakley, & Bauld, 2012; Woods, Sarter, & Billings, 1997), it is important to identify reaction strategies that may and should be used when an unreliable alarm is in the presence of other alarms. Accordingly, the influence of reliability level and the number of additional activated alarms on objective trust, reaction time, and acceptance rate with a 12-alarm array was evaluated using a 2 × 12 split-plot factorial design. Overall a significant linear trend was observed in objective trust measures as the number of additional activated alarms ($p <$ .001). This finding indicates the number of additional activated alarms, instead of the given alarm reliability, was used to calibrate objective trust. Reaction time was found to be quadratic ($p < .001$). Acceptance rate followed a cubic trend ($p < .001$), with significant quadratic ($p = .02$) and significant linear ($p < .001$) derivative trends. This suggests participant response changed from alarm dismissal to acceptance near 50% of

alarm array activation. Finally, there was a significant effect of reliability level ($p < .001$) on acceptance rate, although no significance differences were found between the 50% and 75% groups. Overall, the results constitute evidence for an extension of probability matching theory based on percent system activation and indicate the need to minimize alarms in display design.

This thesis is dedicated to my parents, Wendell and Deborah Allen, whose unwavering support has made this all possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Human beings use automated alarms every day. In the workforce, an automated alarm can signal critical events: warning aircraft pilots to change altitude, aiding doctors as they operate, and helping engineers monitor power plant functions. Often alarm systems enhance human performance; however, factors such as degraded alarm signal, reliability, and trust, can alter how the operator uses automated signaling systems (Breznitz, 1984; Getty, Swets, Pickett, & Gontheir, 1995; Wiegmann, Rich, and Zhang, 2001). Due to the potential consequences of alarm misuse and disuse (Parasuraman & Riley, 1997), a large portion of the alarm literature is dedicated to exploring which factors influence human interaction with individual alarms. Yet, relatively few articles explore how human behavior changes in the presence of multiple simultaneous alarms. Given the increasing use of multiple alarms in environments such as operating rooms and cockpits (Konkani, Oakley, & Bauld, 2012; Woods, Sarter, & Billings, 1997), it is important to identify strategies that may be used when responding to multiple automated alarms. Accordingly, the purpose of this study is to explore alarm response strategies to an unreliable alarm when multiple simultaneous alarms are present.

Alarms

Often the terms alarm, alert, and warning are used interchangeably. However, it is important to distinguish between the three types of signals as they may elicit differing responses. To address the ambiguity of these terms, Bliss and Gilson (1998) defined alarms, alerts, and warnings as part of a taxonomy for emergency signals.

Alarms are signals that require an immediate response from the human operator (Bliss & Gilson, 1998). A common example is a fire alarm. The alarm signals the presence of danger (the fire) and an immediate reaction (evacuation) is required to avoid this danger. *Alerts* signal that a dangerous condition will develop if current conditions continue (Bliss & Gilson, 1998). As such, alerts may not require an immediate response. For example, the gas light indicator signals a condition (low fuel) that will eventually result in danger to the operator (the car shutting off). However, this danger is not currently present, thus, the response does not need to be immediate. *Warnings* are typically written, and indicate that danger may exist given certain conditions (Bliss & Gilson, 1998). A spray paint can contains a warning that the contents are under pressure, and that should external temperatures exceed a specified threshold, combustion may occur.

One of the defining characteristics used to distinguish alarms, alerts, and warnings, is the response required by the operator. Alarms require an immediate response, alerts require an eventual response, and warnings indicate when a response should take place. Additionally, alarms and alerts can be delivered through any modality and are most frequently found as part of an automated system.

Automation

Automated systems complete, or partially complete, a task that could be performed by a human operator (Parasuraman & Riley, 1997). Typically, automation is implemented when a task is too dangerous, difficult, unpleasant, or impossible for a human operator to perform (Wickens, Lee, Liu, & Becker, 2004). For example, the mining industry has begun to use automated mining machines due to the danger mining

presents for human laborers (Lynas & Horberry, 2011). In the power generation industry, trend displays are implemented for use in process control to show the current state of the plant, as well as anticipated states. Monitoring and predicting power plant states may be too mentally demanding for the operator, given the many other tasks they must complete (Moray, 1997). Both of these instances exemplify different *levels of automation* as well as *categories* of automation.

Levels of automation (LOA) are defined by the degree of human involvement, or the level of control, the human operator has over a course of action (Endsley & Kaber, 1999). In contrast, categories refer to the type, rather than level, of automation. Using a simplified version of the information processing model (Figure 1), Parasuraman, Sheridan, and Wickens (2000) proposed four categories of automation: (1) information acquisition, (2) information analysis, (3) decision and action selection, and (4) action implementation.

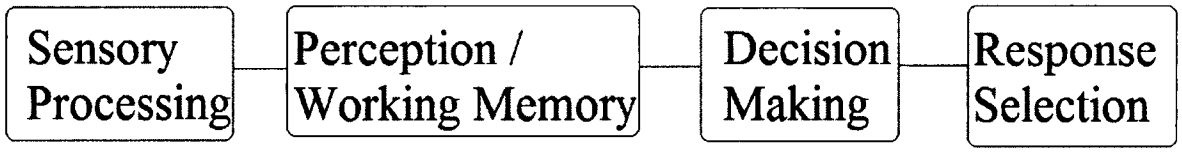

Figure 1. Simplified Model of Human Information Processing System. Adapted from "*A Model for Types and Levels of Human Interaction with Automation*" by R. Parasuraman, T. B. Sheridan, and C. D. Wickens, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30 (3), p. 287. *Copyright 2000 by IEEE.*

Stage 1 of the information-processing model, sensory processing, refers to the sensation and perception of external stimuli by a human being. When applied to automation, this stage corresponds with the first category of automation: information

acquisition, wherein the collection and aggregation of data obtained through sensors is automated.

It is important to note that the level of automation can vary in each category of automation, and different levels of automation can produce differing effects on human performance (Parasuraman, Sheridan, & Wickens, 2000). For example, in the information acquisition category, sensors that automatically adjust their positions to optimize data acquisition would characterize low levels of automation. The human operator must manually sort and prioritize the data. The highlighting of important data by the automation would represent a higher LOA. An even higher LOA would consist of filtering information, in which the automation reviews the data and displays only certain information to the operator, resulting in less human involvement. As a result, it is essential to specify the level and category of automation under investigation to ensure proper generalization of results. This is especially important for theoretical predictions, as the predicted outcomes may apply to only certain types or levels of automation. For the purposes of this study, multiple alarm signal response will be examined by using a sensor-based signaling system, which represents low levels of automation within the first (information acquisition) category of automation.

Sensor-Based Signaling Systems

The term "sensor-based signaling systems" (henceforth called signaling systems) was created to describe automated systems used to monitor sources of potential hazards and to direct user attention as needed (Bliss & Gilson, 1998). From a theoretical standpoint, signaling systems correspond to the first stage of information processing, sensory processing, due to the automation's purpose of gathering external data. This

purpose is similar to how a human would sense and process external stimuli. Using the corresponding categories proposed by Parasuraman et al. (2000), signaling systems are thus categorized as information acquisition automation.

A common example of a signaling system is a smoke detector. Once a threshold of smoke concentration has been reached, the system will direct human attention to the potential threat of fire. Although this example demonstrates the use of a signaling system for an alarm, these systems can also be implemented to issue alerts or warnings (as previously defined), thus providing a broad range of applications (Bliss & Gilson, 1998; Meyer 2004). Given this range of functions, signaling systems are perhaps one of the most familiar forms of automation; signaling systems can be found in security monitoring, aviation, medicine, transportation, power generation, and military application domains (Wickens, 2004).

Trust

Due to the prevalence of signaling systems, it is vital to understand the factors affecting human-automation interaction. If the human operator does not respond appropriately to a signaling system, then the value of the automation is diminished. In critical situations, an inappropriate response can even result in death. For example, a smoke detector may signal the possible presence of fire, yet people may ignore the signal and fail to evacuate.

Because of the potential consequences associated with ignoring a sensor-based signal, considerable research has been devoted to identifying key factors that influence an operator's decision to dismiss or ignore critical signals. One of the most prevalent factors thought to influence human-automation interaction is trust. Simply, if the operator does

not trust an automated system, such as a smoke detector, he or she is less likely to use that system. This relationship between trust and automation use has been the subject of considerable research (Lee & Moray, 1994; Muir & Moray, 1996; Lee & See, 2004; Muir, 1986, 1994; Rice, 2009, Wiegmann et al., 2001).

Subjective measures of trust. In exploring the construct of trust, Muir (1987) suggested that trust in automation is similar to interpersonal trust. Under this theory, trust in automation can be affected by the same factors that affect trust in humans. For example, Rempel, Holmes, and Zanna (1985) identified three dimensions of interpersonal trust: predictability, dependability, and faith. These dimensions are comparable to human-automation dimensions of trust suggested by Sheridan (1981): reliability, dependence, familiarity, and robustness.

Similarly, Mooreman, Deshpande, and Zaltman (1993) defined interpersonal trust as a "willingness to rely on an exchange partner in whom one has confidence" (p. 82), indicating reliance and confidence as key components of interpersonal trust. Wiegmann, et al. (2001) built upon this and defined subjective measures of automation trust as users' confidence ratings and verbal estimates of reliability. Although the nature of trust is still debated in the literature, many constructs of trust include dimensions of reliability, confidence, and/or dependability (Jian, Bisantz, & Drury, 2000; Rempel et al., 1985; Sheridan, 1981; Wiegmann et al., 2001). However, popular measures of subjective trust, such as the one by Jian et al. (2001), have yet to be validated. Thus the use of subjective measures of trust can be controversial.

Objective measures of trust. Trust has also been measured using objective (behavioral) measures. Muir and Murray (1996) found a positive correlation between

trust and the amount of control allocated to the system by operators using a virtual pasteurizer plant. Similarly, trust has been found to be a factor in reliance on augmented vision system in target identification (Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001). Field studies have also identified the role of trust through observations of autopilot use and flight management systems (Moiser, Skitka, & Kore, 1994).

There is some evidence that reaction time may be a particularly useful indicator of trust. A qualitative study by Getty, Swets, Pickett, and Gonthier (1995), found participants reacted more quickly to alarms with high Positive Predictive Values than alarms with low Positive Predictive Values. Subsequently, some authors choose to measure trust using reaction time (Rice, 2009). Rice wrote, "I assumed that when participants trusted the automation, they would quickly agree with the aid" (p. 312). Although Rice (2009) does not cite the reasons for his assumptions, they may be based on the earlier work by Getty, Swets, Pickett, and Gonthier.

Conversely, there is also evidence that trust does not mediate the relationship between reliability and reaction time. Chancey (2013) empirically assessed how subjective measures of trust mediate response behavior. It was found that trust partially mediated the relationship between reliability rate and agreement rate, however, trust did not mediate the relationship between reliability and reaction time. These findings were supported by a subsequent study in which the role of trust as a mediator for reliability and reaction time was analyzed (Chancey, Proaps, Bliss, 2013). Additionally, Wiegmann (2001) did not find consistent correlations between subjective measures of trust and reaction time, suggesting that reaction time may not be a good indication of trust.

It is possible that the decreased reaction times observed in Rice's work are due to participants' confidence in their own responses, and not the participants' trust in the alarm itself. It is conceivable that more highly reliable automation would induce higher levels of response confidence in participants, and thus reaction time may be a function of response confidence and not an indication of trust in the alarm. As a result of the seemingly conflicting evidence, the relationship between trust and objective measures is also controversial in the literature.

In an effort to more fully understand the role of trust in multiple alarm situations, both subjective and objectives measures of trust are used in this study.

Reliability

Because no automation can ever be 100% reliable, unreliability is an inherent problem with all automation. In keeping with the smoke detector example, the smoke detector's sensor collects data about particles in the air. These data are processed using an algorithm to determine if the preset threshold has been met, at which point the smoke detector will signal the presence of smoke. However, if the threshold is too liberal, it will signal the presence of smoke when relatively few particles are in the air, which might be indicative of dust accumulation or a slight wisp of smoke from extinguishing a candle. This could constitute a false alarm, depending on the consequences associated with the presence of smoke. Conversely, if the threshold is set too conservatively, the smoke detector may fail to signal the human operator, despite the presence of smoke. This constitutes a miss.

Both false alarms and misses have been studied extensively in the literature. Evidence suggests false-alarm and miss-prone systems may evoke differing responses

from the human operator (Parasurman & Riley, 1997). In false-alarm prone systems, the operator may not trust that the alarm is a true alarm due to the high occurrence of false alarms. Consequently, operators react more slowly (Getty et al., 1995), ignore, or disable the alarm (Sorkin, 1988). This response behavior has been termed the "cry-wolf" effect, based on Breznitz's (1984) work examining behavioral and physiological responses to false alarms. Similarly, specific patterns of behavior have been associated with miss-prone alarms. In miss-prone systems, operators may develop a maladaptive automation reliance behavior called misuse, in which the operator fails to detect a miss due to an over-reliance on the system to detect all hazards (Parasuraman & Riley, 1997). The operator trusts the automation to accurately detect and identify all hazards.

Because the unreliability of the signaling system affects operator trust (Lee & See, 2004; Meyer, 2001; Rice, 2009), it is important to consider alarm reliability when investigating operator trust and response behaviors. In a study by Wiegmann et al. (2001), higher reliability levels resulted in higher agreement rates, quicker decision times with affirmative decisions, higher confidence ratings, and higher subjective ratings of automation reliability. It was also found that operators were sensitive to changes in reliability (Wiegmann et al., 2001).

In some cases, lower reliability levels of automation can be so detrimental to performance it would be better if there were no automation at all (Wickens & Dixon, 2007). The level at which performance falls below baseline (performance levels with no automation) is estimated to be at 70% reliability (Wickens & Dixon, 2007). This estimate was determined using regression analysis of the results of over 40 studies. These studies included Type 1 (Information Acquisition) and Type 2 (Information

Analysis) automation, miss-prone and false alarm-prone systems, as well as a variety of opaque and clear systems. It should also be noted that reliability had a more pronounced effect on performance when workload was high, such as a dual task paradigm. Given the variety of conditions represented in the data set used, 70% can only be used as a general estimate, and a range of reliabilities should be used when possible.

Multiple Alarms

The preponderance of literature is dedicated to investigating single automated sensor-based signals. However, relatively little research has identified strategies for responding to multiple contiguous sensor-based signals. This is a critical omission in the literature because technology has afforded the development of increasingly complex systems, which can often have more than one potential hazard, suggesting the need for multiple sensor-based signals. To illustrate, airplane cockpits and nuclear power plants can have potentially hundreds of alarm signals. The likelihood of needing to resolve multiple signals co-located in the same environment can be high. In an extreme case, during the Three Mile Island nuclear power incident, more than 500 annunciators changed status (Sheridan, 1981). Moreover, confusion over the relationship between ambiguous indicators can pose a problem to operators who may be forced to make a decision (Gilson, Mouloua, Graft, & McDonald, 2001).

Several strategies for single alarm response have been previously observed in the literature. For example, Bliss, Gilson, and Deaton (1995) found evidence of probability matching behavior, as well as what the authors termed an "optimal strategy" response pattern. Approximately 10% of the participants adopted the "optimal strategy" and became "extreme responders" such that in accepting a 75% reliable alarm at every

presentation, participants ensured that they were correct 75% of the time (1.00 × .75= .75). In contrast, the majority of participants used a probability matching strategy, which results in lower overall accuracy.

Probability matching is a strategy in which in alarm acceptance is calibrated based on alarm reliability. Statistically, if a participant accepts a 75% reliable alarm in only 75% of presentations, then the resulting correct alarm acceptance rate would be 56.25% (0.75 × 0.75= 56.25). Given that various alarm response strategies have been identified with single alarms, it is likely that an operator may also use one or more strategies when responding to multiple alarms.

Two strategies proposed by Keller and Rice (2009) are component-specific trust and system-wide trust. In component-specific trust, an unreliable signal is viewed as an individual component that is separate from the other sensor-based signals that may also be present. Consequently, acceptance rates of the other, more reliable, alarm signals in an array should be unaffected by a single unreliable alarm signal. Alternatively, the operator may adopt system-wide trust, in which the reliability of an alarm signal is evaluated based on the entire system of sensor-based signals. Keller and Rice evaluated these two theories through a series of studies.

In an initial study, Keller and Rice (2009) presented participants with two gauges, the second of which was always 100% reliable. The other gauge was 70%, 85%, or 100% reliable, depending on group assignment. Sensitivity of the second (100% reliable) alarm decreased in conditions where the first alarm was 70% or 85% reliable. Thus, the imperfect alarm impacted sensitivity for the always 100% reliable alarm. This "dragging down" effect was later observed with alarm agreement rates using a larger eight-alarm

array (Geels-Blair, Rice, & Schwark, 2013). Agreement rates for the always 100% reliable alarms (alarms 2-8) lowered in conditions where the first alarm was imperfect. The effect on seemingly unrelated signals, in a variety of signal array sizes, suggests operators adopted system-wide trust as opposed to component-wide trust.

It should be noted that all measures of trust in these two studies were objective: reaction time and alarm acceptance. Some researchers, such as Wiegmann et al. (2001), recommend that subjective measures should be used to indicate trust (a psychological construct), and objective behavioral measures should to indicate automation reliance. This recommendation is based on findings that objective and subjective measures of trust were inconsistently correlated (Wiegmann et al., 2001).

Additionally the signals used by Rice et al. were not completely opaque, the operator could verify the accuracy of the alarm by comparing the gauge value to a given safe value; however, it required significant cognitive resources to verify alarm accuracy due to their complexity. There are also issues concerning when the alarms were active. In the 2009 experiment by Keller and Rice, only a single alarm was activated at any given time. As previously mentioned, it is possible, and in some environments likely, that multiple signals will simultaneously indicate a hazard. Although it was feasible for more than one alarm to be activated in each trial of the 2013 study (Geels-Blair, Rice, & Schwark, 2013), the results were not analyzed based on the number of activated alarms present. The work by Gilson, Mouloua, Graft and MacDonald (2001) addresses some of these issues by examining confidence when multiple-alarm signals are present.

In a series of studies by Gilson et al., participants were given an array of six alarms, one of which was marked "test" alarm. Participants were told that the test alarm

was only 50% reliable, meaning that it was actually a false alarm half of the time the alarm was activated. A series of trials were then presented in which additional alarms were activated with the test alarm. When only the test alarm was activated, participant's average confidence that the test alarm was a true alarm was 23%, significantly lower than the given 50% reliability level. As the number of additional active alarms increased, so did participant confidence. An activation of all six alarms produced an average confidence rating of 97%. Additionally, the increase of confidence level with additional alarm activation produced a significant linear trend. Gilson et al. (2001) subsequently postulated that confidence level is founded upon the overall number of activated alarms. These changes in confidence level suggest that participants evaluate spatially contiguous alarms as part of a larger system and not independently, similar to the findings of Geels-Blair, Rice, and Schwark (2013).

Gilson et al.'s (2001) research raises many intriguing questions; however there are several things to note concerning his work. First is the issue of reliability. As previously mentioned, the work by Wickens and Dixon (2007) recommends that a variety of reliability levels should be used, with 70% as the possible threshold for automation related performance increases. A 50% reliability level is akin to guessing and may not be the most ecologically valid reliability level. Additionally, the context of the alarm signal was not given to the participant. Although such situations may exist in the real world, for example someone may hear a smoke alarm while in a different room from the fire, the lack of alarm context in this experiment raises the question of ecological validity.

Furthermore, Gilson et al. (2001) used six-alarm arrays in their studies. If participants are indeed basing confidence on the percentage of the overall all system

activation, this would necessitate only six possible percentage estimates. Four of these six possible activation estimates would occur when the number of activated alarms, divided by the total number of alarms, results in commonly used fractions such as 1/3, 1/2, 2/3, and 1/1. If a larger array had been used, a greater number of percentage estimations formed from uncommon and more difficult to evaluate fractions would be required. Research suggests that people tend to underestimate high probabilities and under estimate low probabilities (Hollands & Dryer, 2000). It is possible that the increased complexity of a larger array may reveal a non-linear pattern of estimation, similar to that of previous research on proportion estimation. If Gilson's idea of system percent-activation is correct, then an array with one out of six alarms activated should produce the same percent confidence rating as an array with two out of 12 alarms activated.

Most importantly, although confidence has been identified as a critical component of automation-human interaction (Jian, Bisantz, & Drury, 2000; Rempel et al.1985; Sheridan, 1981; Wiegmann et al., 2001), confidence estimates alone do not fully capture the construct of trust. A more robust measure of trust and its different dimensions, to include the dimension of confidence, would allow for a more accurate interpretation of the results.

Goal of Present Study

The purpose of the present study was to evaluate the system-percentage strategy proposed by Gilson et al. (2001) using a larger, 12-alarm array with three levels of reliability. Additionally, participants were required to accept or dismiss the alarm and answer a trust questionnaire, in order to obtain both subjective and objective measures of

trust. The alarms were opaque and false alarm prone; however, a context was given to participants to provide ecological validity.

Based on the previous literature discussed, there are four hypotheses. First, a greater number of activated alarms will lead to higher acceptance rates on an unreliable test alarm than when fewer alarms are activated. This hypothesis is based on the work by Keller and Rice (2009), and Geels-Blair, Rice, and Schwark (2013), where it was found reliability of surrounding alarms (as perceived by the number of alarm activations over time) were a factor in alarm acceptance by the participant. Given the previous influence of surrounding alarms, it is anticipated that participants in this study will likewise use the activation of surrounding alarms when responding to the test alarm, resulting in a higher acceptance rate when a greater percentage of the display is activated.

Second, subjective trust of the unreliable test alarm will increase as the number of additional activated alarms increases. Gilson et al., (2001) found confidence, a dimension of subjective trust, increased as the number of additional activated alarms increased. Consequently, it is expected that subjective measures of trust will also increase as the number of additional alarms increases.

Third, reaction time is hypothesized to follow a quadratic trend: as the number of activated alarms reaches the extremes (all or none of the array), participants will respond more quickly to the unreliable test alarm. When studying multiple alarms, Gilson et. al (2001) measured participants' confidence in the test alarm. Gilson found when all of the alarm array was activated, participants were confident the alarm was true. When no additional alarms were activated, participants expressed low confidence that the alarm was true, suggesting that the participants were confident that it was a false alarm. As

previously discussed, it is possible that the decreased reaction times observed in previous work (Keller & Rice, 2009; Geels-Blair, Rice, and Schwark 2013; Rice, 2009) may be due to participants' confidence in their own responses, and not the participants' trust in the alarm itself. If reaction time is a function of response confidence, it can be hypothesized that participants will have faster reaction times at higher confidence levels. Thus, based on this previously found confidence pattern (Gilson et al., 2001), it is predicted there will be higher confidence in a response (accept or dismiss) at the extremes of multiple alarm activation (all and none), producing an overall quadratic trend in reaction time.

Finally, it is hypothesized that there will be an anchoring effect: participants exposed to highly reliable alarms will indicate significantly higher mean trust ratings compared to participants exposed to less reliable alarms. Many researchers have discussed the role of reliability, however Wiegmann, et al., (2001) explicitly examined how different levels of reliability effect both subjective and objective trust, finding participants to be sensitive to differing levels of reliability. Accordingly, it is predicted that participants in this study will exhibit higher levels of subjective trust at the 100% reliability level than the 75% and 50% level, respectively. Similarly, there will be higher levels of subjective trust at the 75% reliability level than the 50% reliability level.

CHAPTER II

METHOD

Research Design

This experiment employed a 2 × 12 mixed factorial experimental design. The between-groups independent variable, alarm reliability, consisted of three levels: 50%, 75%, and 100% true alarms. The within-groups independent variable was the number of additional activated alarms. Dependent variables were reaction time, subjective trust, and alarm acceptance rate.

Participants

A power analysis conducted using G*Power 3.1 software estimated a sample size of 42 participants for this study. Due to the lack of established effect sizes with multiple alarms, a conservative small to medium effect size was used to establish the target sample size for the primary ANOVA analysis (.15, $\alpha = .05$ and $\beta = .80$). Based on this power analysis, 44 participants (19 males and 25 females) were recruited from the Old Dominion University Psychology Department subject pool and compensated with research participation credit. The mean age of participants was 20 years ($SD = 4.2$, min = 18, max = 45). All participants had normal or corrected-to-normal vision and did not report any sensorimotor deficits. On average, participants reported playing video games 3.0 hours per week ($SD = 2.7$, min = 0, max = 7).

This study was approved by Old Dominion University Institutional Review Board (IRB). Signed informed consent was obtained from each participant prior to beginning the experiment.

Apparatus

All experimental equipment and programs were controlled by a desktop computer with an Intel Core i7 2.67 GHz processor and 9.00 GB system RAM. Visual stimuli were presented on a standard 22-inch LG LCD color computer monitor using SuperLab 4.5.2 software. The computer monitor was approximately 12 cm above the surface of the desk and 60 cm from the seated participant.

Stimuli. Participants were presented with a 12-alarm array, as illustrated in Figure 2.

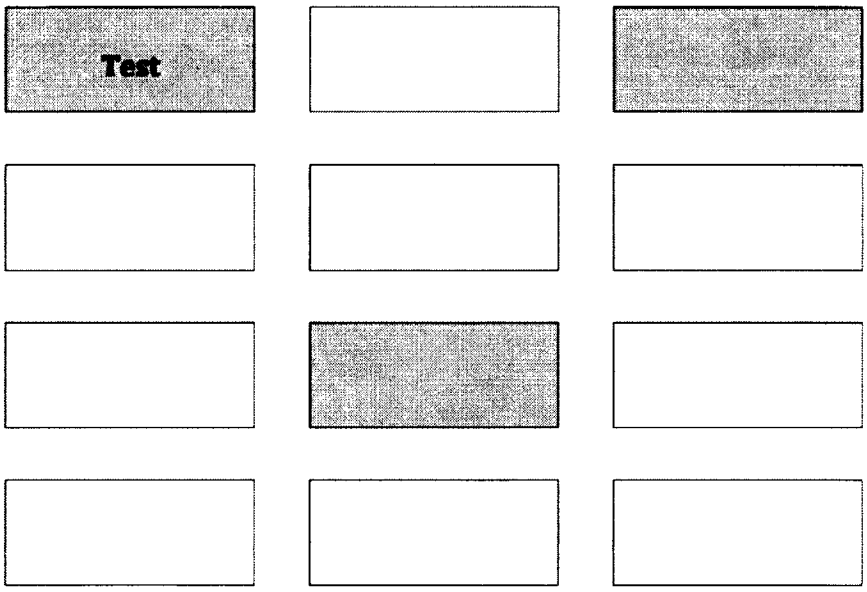

Figure 2. Sample 12-Alarm Array with Three Activated Alarms.

The alarm stimulus consisted of four rows of three boxes, creating a 12-box array. The boxes were 6 cm × 3 cm with a 1 cm space separating all boxes. Based upon the placement of the monitor and typical viewing distance, each box subtended average

viewing angles of 5.8 degrees horizontally and 2.9 degrees vertically. Each box represented an alarm, and activated alarms were displayed in gray. Prior research has found the perception of red enhances motor function response as compared to lightness-matched gray alarms (Elliot & Aarts, 2011). Although red is often used in alarm signals, grey was chosen for the stimuli in this experiment so as not to preclude color deficient or color blind students. The upper left alarm was labeled "test" in all presentations of the stimulus, to indicate the unreliable alarm the participant would be responding to during the experiment.

Response Method. A Cedrus model RB-530 response box was used to record responses. The response box contains a subprocessor for low latency and is accurate to within one millisecond. The RB-530 buttons are approximately 2.0 × 2.4 cm in size and located 1.0 cm apart. "Accept" and "Dismiss" labels were affixed to the left and right response button, respectively. A standard keyboard number pad was used to record participant responses to a trust questionnaire.

Tasks and Measures

Primary Task: Multi-Attribute Task Battery (MATB II). The MATB II program simulates the kinds of tasks that pilots perform during flight (Santiago-Espada, Meyer, Latorella, & Comstock, 2011). Participants were asked to perform the compensatory tracking task available in this battery (see Figure 3).

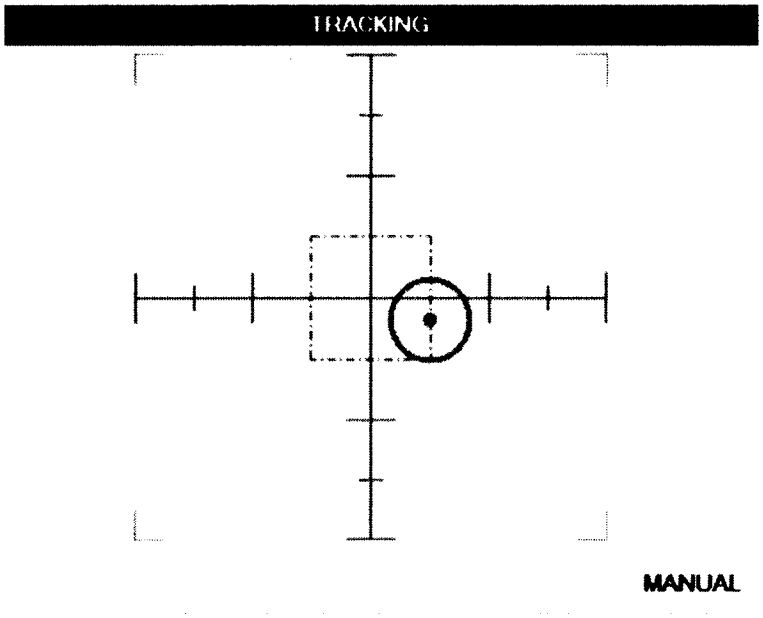

Figure 3. Compensatory Tracking Task.

Using a joystick, participants were asked to keep a blue reticle as close to the center of the pair of crosshairs as possible. This center location is further identified by a square surrounding the area. During the task, the reticle continuously drifts in random directions. Participants must make compensatory movements to keep the reticle centered on the crosshairs. The purpose of the task is to simulate maintaining level flight while environmental factors, such as wind, affect the aircraft. The root mean square error of the reticle was recorded every 15 seconds. The root mean square error is considered an indication of the stability of tracking performance, as it describes the error of the participant in holding the reticle at center.

Secondary Task: 12-Alarm array configuration. Participants were presented with an array configuration consisting of 12 alarms. The top left alarm was designated and labeled the "test" alarm. Typically, a miss is generated when an alarm fails to signal when there is a real hazard. In this experiment, the test alarm was always activated, thus

the test alarm could only produce real or false alarms. On any given trial, from one and eleven additional alarms were also activated. Presentation of an alarm configuration occurred after a random inter-stimulus-interval of 8, 12, or 16 seconds to prevent participants from forming a response rhythm. The positions of the additional activated alarms were pre-selected at random. Participants were prompted to accept or dismiss the test alarm, using the Cedrus model RB-530 response box, as quickly and accurately as possible. Reaction time was measured in milliseconds from the presentation of the stimulus to response input (alarm acceptance or dismissal). Acceptance rates for each alarm configuration were calculated by dividing the number of alarm acceptance responses by the total number of presentations. No response feedback was given.

Trust Questionnaire. A modified version of Jian, Bisantz, and Drury's (2000) human-automation trust questionnaire was used to assess participants' trust in the test alarm. The original survey consists of five items assessing operator distrust, and six items assessing operator trust. As participants were presented with the trust questionnaire after every trial, the complete Jian, Bisantz, and Drury trust questionnaire would have been potentially fatiguing to participants. Thus, the modified version used in this study retained questions only from the trust portion of the questionnaire, which best aligned with previous research on dimensions of trust (Muir, 1987; Rempel et al., 1985; Sheridan, 1998). Additionally, when examining the human-automation trust questionnaire, the trust items were compared to dimensions of trust previously identified, most notably by Sheridan (1988; familiarity, reliability, and confidence) and by Muir (1987; faith, predictability, and dependability). Two trust questionnaire items from the Jian, Bisantz, and Drury measure were unrelated to previously identified trust dimensions (i.e., this

system provides security and this system has integrity). These two questions were also removed. The resulting questionnaire consisted of four questions:

- How confident are you the test alarm is a true alarm?
- How much do you trust the alarm?
- How reliable is the alarm?
- How dependable is the alarm?

Participants were instructed to respond honestly to all questions using a scale of 0-100. As an example, participants were told a response of 100 to the question "how confident are you the test alarm is a true alarm?" indicates full confidence, and zero indicates no confidence. Responses were entered using a standard numeric keypad. The four questions were displayed in random order after each alarm configuration presentation. This randomization prevents survey bias based on question order. Additionally, each alarm configuration was randomly presented a total of five times throughout the experiment.

Separate dimension scores were calculated by averaging the responses to each of the four questions presented for each alarm configuration. A composite trust score was then computed by summing all of a configuration's dimension scores. As there is no theoretical justification to weighing one trust dimension greater than another, the approach of using unweighted averages and summations was used. Composite trust scores could range from 0-400.

Procedure

Participants were given an overview of the experiment and written informed consent was obtained (see Appendix A). A brief medical questionnaire (see Appendix B)

was used to screen for sensory or motor deficits. Any sensorimotor deficits would result in exclusion from the study. Next, participants were randomly assigned to one of three different reliability groups (50%, 75%, and 100%) and seated in front of a standard desktop computer.

A vignette was given to participants instructing them to pretend they were an airplane pilot. Participants were told they were in charge of flying the plane, their primary task, as well as responding to a panel of alarms, their secondary task. An example panel of a random 12-alarm array configuration was then displayed on the right-most screen. Participants were informed each gray box represented an activated alarm indicating something was wrong with the plane. Depending on group assignment, participants were told the test alarm in each configuration was true 50%, 75%, or 100% of the time. As the pilot, participants were told they were responsible for either accepting the test alarm as a true alarm or dismissing the alarm as a false alarm. The "accept" and "dismiss" alarm response buttons were then pointed out to the participant. Participants were asked to respond as quickly and accurately as possible, and were reminded of the real-world consequences associated with alarm acceptance and dismissal: accepting an alarm as true would alert flight control and possibly delay or ground the flight, something the pilot should avoid if the alarm is not true. Alternatively, the safety of the passengers is also the responsibility of the pilot, and ignoring a true alarm may endanger the passengers onboard. No information was given concerning the reliability or relatedness of the other alarms in the panel. Once the alarm was accepted or dismissed participants were prompted to respond to the trust questionnaire. The numeric keypad was then demonstrated for questionnaire response.

Following the secondary alarm task demonstration, the MATB-II tracking task was introduced to participants of the left-most screen. After familiarizing themselves with the tracking task for approximately 2 minutes, participants practiced both the primary and secondary task together for three randomly chosen alarm configurations of the secondary task. Completing both tasks on separate screens required a division of attention by the participants.

Once the practice session was completed, participants were given the opportunity to ask any questions before the start of the experiment. Participants were presented with five instances of each 12-alarm array configurations, in random order, resulting in 60 trials. At the conclusion of the experiment, the participants were debriefed and dismissed. The experiment lasted approximately 40 minutes.

CHAPTER III

RESULTS

All statistical tests were conducted using PASW Statistics 20 software, with α = .05. No family-wise alpha corrections were made as hypotheses were *a priori* .

It should be noted that an alpha of level represents the probability of a Type I error, or detecting a relationship between variables when there is not one. Conversely, a Type II error represents the probability of failing to detect a relationship when there is one. A higher alpha level results in a higher probability of a Type I error, but a lower probability of a Type II error. The balance of a Type 1 and Type II, and the associated consequences of each, should be taken into account when choosing an alpha level for an experiment. However, what is considered an acceptable alpha level, and thus the best balance of Type I and Type II errors, is conventionally set within a discipline (Maxwell & Delaney, 2004, p. 24), such as the α =.05 used in this study.

However, a conventionally set alpha level still does not fully address the concerns of a Type II error when examining a specific study within a discipline. Given this issue, Maxwell and Delaney (2004, p. 24) suggest that power, in addition to the set alpha level, should be taken into consideration when evaluating the validity of a statistical conclusion. Power is the probability of rejecting the null hypothesis when it is false. Statistically it is equivalent to 1 – the probability of a Type II error. Higher power thus corresponds with a lower likelihood of a Type II error. All statistical tests performed in this experiment achieved an observed power level greater than .80.

To address the statistical assumption for normality, histograms of the data were visually inspected for unimodal distribution. Additionally, a skew and kurtosis threshold

of |2| was used, as per the recommendations of Maxwell and Delaney (2004, p. 115). Levene's Test was used to assess homogeneity of variance for the between-subject variable (reliability level). These assumptions of normality and homogeneity were generally met, and ANOVA is robust to violations of normality and moderate violations of homogeneity (Maxwell & Delaney, 2004, p. 110). For the within-subject variable (additional alarm activation), Mauchley's tests were conducted to assess sphericity. The assumption of sphericity was violated and a Greenhouse-Geisser correction was used in all cases. For a detailed report of descriptive analyses of the data, see Appendix C.

The 50% and 75% reliability level groups each contained 15 participants and the 100% reliability level group contained 14 participants. However, reaction time data for one participant in the 50% reliability group was removed because the participant left the room during the experiment. Four participants in the 100% reliability group and one participant in the 50% reliability group adopted an optimization strategy wherein the participant accepted the alarm at every presentation. This behavior of extreme responding has been previously observed (Bliss, Gilson, & Deaton, 1995), and suggests that some operators may use alternative strategies. Nevertheless, these types of responders would also be present in real-world alarm scenarios, and their inclusion in this study increases ecological validity. Thus, these participants' data were not adjusted or removed from the analyses. No additional outliers were removed from the data set for similar reasons of ecological validity.

Finally, polynomial trend analyses produce a-1 trend components, making an 11^{th} order polynomial trend possible in this experiment. These higher order trends are often uninterruptable. Additionally, high numbers of polynomial trends increase the likelihood

of a Type 1 error as well as present a danger of over fitting the data. However, data may represent combination of several pure polynomial trends, and thus it is recommended that higher order trends should be tested (Maxwell & Delaney, 2004, p. 259). As and example Maxwell and Delaney point out a negatively accelerated curve would increase as X increases, but the increases themselves become smaller over time (2004, p. 259). This example represents a model with both linear and quadratic components. It is thus important to report some higher order polynomial trends, as well as the pure polynomial trend that may best account for the data. For this experiment, a visual inspection of the graphed data and effect size are taken into account before reporting the highest interpretable significant polynomial trend in the text.

Hypothesis One: Acceptance Rate

The first hypothesis, that a greater number of activated alarms will result in significantly higher alarm acceptance rates of an unreliable alarm, was tested using a 2 × 12 split-plot ANOVA. The between-groups independent variable was reliability level (50% and 75%), the within-groups independent variable was the number of additional activated alarms (0-11), and the dependent variable was acceptance rate. The 100% reliability was group not included in this analysis as it did not represent an unreliable alarm, and thus did not represent an alarm that would be present in the real world.

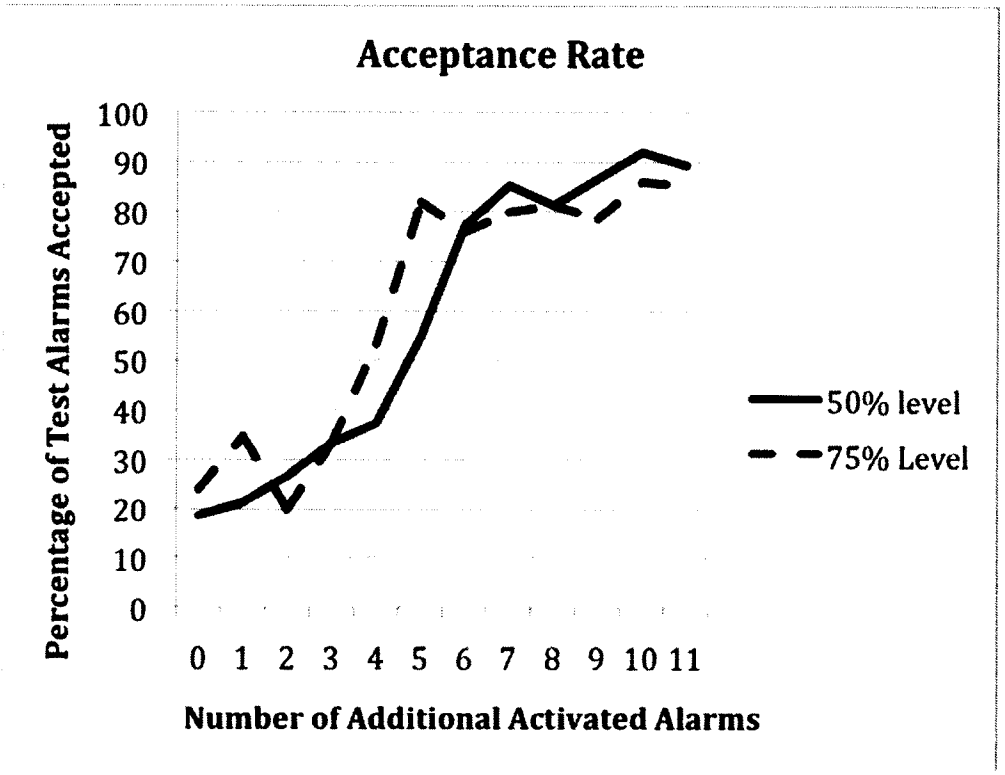

Figure 4. Acceptance Rate as a Function of Number of Additional Activated Alarms.

There was no significant interaction between additional alarms and reliability level; however, a significant main effect of additional alarms was found, $F(2.65, 74.06) = 27.71$, $p < .001$, $\eta_p^2 = .50$ (see Table 1). An *a priori* polynomial trend analysis of the main effect of the number of additional alarms revealed a significant cubic trend, $F(1, 28) = 11.05$, $p < .001$, $\eta_p^2 = .28$, a significant quadratic trend, $F(1, 28) = 6.65$, $p = .02$, $\eta_p^2 = .19$, and a significant linear trend, $F(1, 28) = 51.62$, $p = <.001$, $\eta_p^2 = .65$. There was no significant difference between reliability groups.

Table 1

Split-plot ANOVA for Effects of Number of Additional Alarms and Reliability Level on Acceptance Rate

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Reliability Level	0.06	1.00	0.06	0.27	0.61	0.01
Error (between)	6.16	28.00	0.22			
Additional Alarms	24.62	2.65	9.31	27.71	<.001	0.50
Linear	22.06	1.00	22.06	51.62	<.001	0.65
Quadratic	0.75	1.00	0.75	6.65	0.02	0.19
Cubic	0.87	1.00	0.87	11.05	<.001	0.28
Additional Alarms x Reliability Level	0.99	2.65	0.38	1.12	0.34	0.04

Note. This table displays the omnibus sources of variance, as well as follow-up polynomial trend analyses of the main effect of the number of additional alarms.

Hypothesis Two: Subjective Trust. A 2 × 12 split-plot ANOVA was used to test the second hypothesis, that trust in and unreliable test alarm would increase as the number of additional activated alarms increases. Similar to the first hypothesis, the 100% reliability group was not included in this analysis, as it did not represent an unreliable alarm. The independent variables were alarm reliability (between-groups; 50% and 75%) and the number of additional active alarms (within-groups; 0-11).

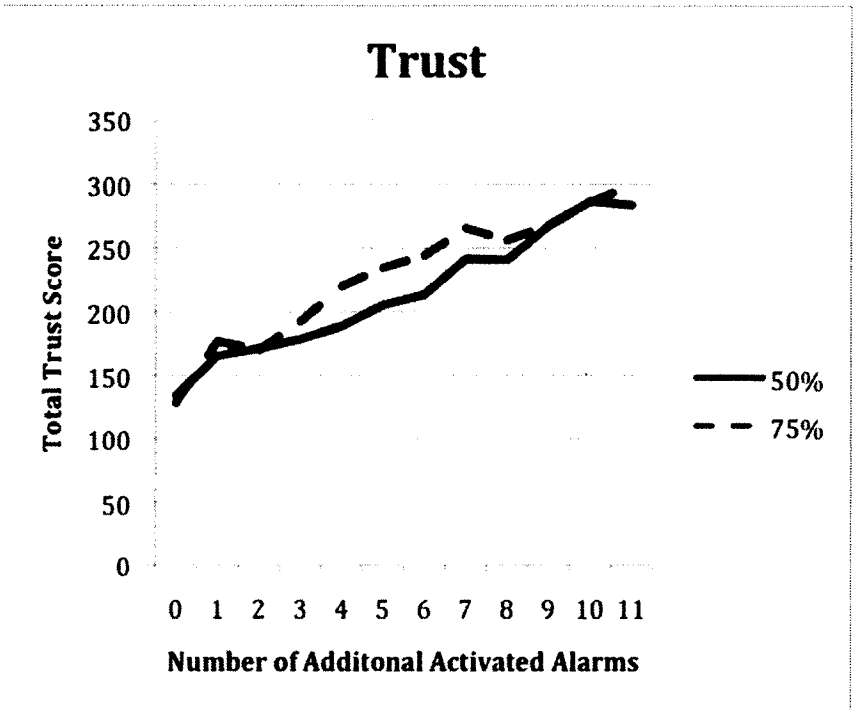

Figure 5. Composite Trust Score as a Function of Number of Additional Activated Alarms.

A significant main effect of additional alarms was found, $F(1.73, 48.41) = 22.33$, $p < .001$, $\eta_p^2 = .44$ (see Figure 5 and Table 2). An *a priori* polynomial trend analysis of the main effect of the number of additional alarms revealed a significant linear trend, $F(1, 28) = 29.68$, $p = <.001$, $\eta_p^2 = .52$. There was no significant difference in trust between the reliability levels. Additionally, there was no significant interaction between additional alarms and reliability level.

Table 2

Split-plot ANOVA for Effects of Number of Alarms and Reliability Level on Composite Trust Score

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Reliability Level	16922.71	1.00	16922.71	2.74	0.11	0.09
Error (between)	173269.68	28.00	6188.20			
Additional Alarms	835700.55	1.73	483339.02	22.33	<.001	0.44
Linear	817263.62	1.00	817263.62	29.68	<.001	0.52
Quadratic	3748.39	1.00	3748.39	2.05	0.16	0.07
Cubic	647.64	1.00	647.64	0.80	0.38	0.03
Additional Alarms x Reliability Level	14540.71	1.73	8409.82	0.39	0.65	0.01

Note. This table displays the omnibus sources of variance, as well as follow-up polynomial trend analyses of the main effect of the number of additional alarms.

Hypothesis Three: Reaction Time

The third hypothesis, that reaction time will produce a quadratic trend, was evaluated using a 2 × 12 split-plot ANOVA. The 100% reliability group not included in this analysis as it did not represent an unreliable alarm. The independent variables were alarm reliability (between-groups; 50% and 75%) and the number of additional active alarms (within-groups; 0-11).

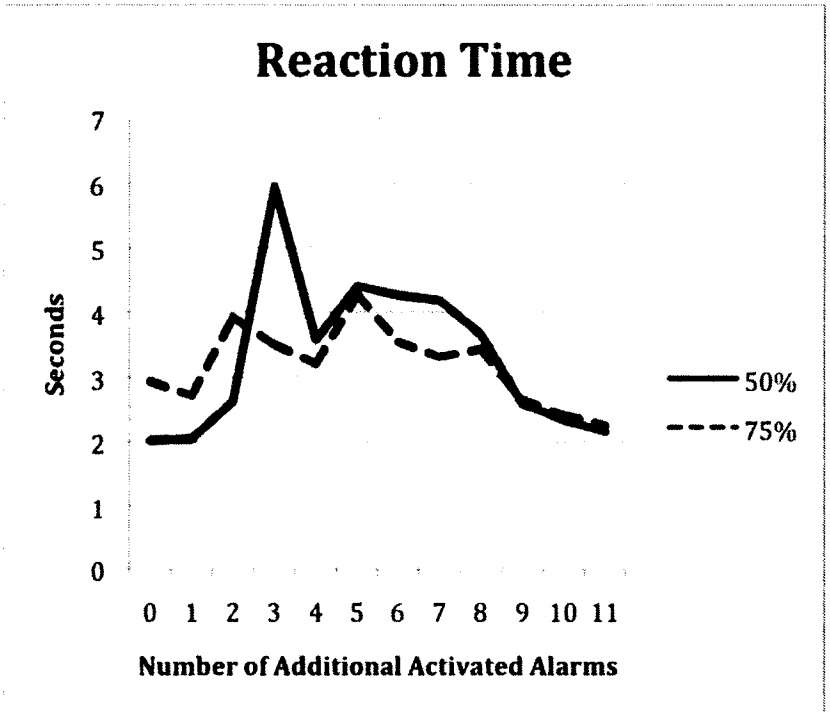

Figure 6. Reaction Time as a Function of Number of Additional Activated Alarms.

There was no significant interaction between additional alarms and reliability level; however, a significant main effect of additional alarms was found, $F(3.05, 82.39) = 4.33$, $p = .007$, $\eta_p^2 = .14$ (see Figure 6 and Table 3). An *a priori* polynomial trend analysis of the main effect of the number of additional alarms revealed a significant quadratic trend, $F(1, 27) = 27.58$, $p = <.001$, $\eta_p^2 = .51$. There was no significant difference between the reliability levels.

Table 3
Split-plot ANOVA for Effects of Number of Additional Alarms and Reliability Level on Reaction Time

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Reliability Level	1666896.21	1.00	1666896.21	0.04	0.85	0.00
Error (between)	1292463031.39	27.00	47869001.16			
Additional Alarms	224003432.25	3.05	73406124.03	4.33	0.01	0.14
Linear	8330236.17	1.00	8330236.17	2.70	0.11	0.09
Quadratic	155343065.94	1.00	155343065.94	27.58	<.001	0.51
Cubic	5206592.92	1.00	5206592.92	0.85	0.36	0.03
Additional Alarms x Reliability Level	73294309.71	3.05	24018610.50	1.42	0.24	0.05

Note. This table displays the omnibus sources of variance, as well as follow-up polynomial trend analyses of the main effect of the number of additional alarms.

Hypothesis Four: Anchoring Effect

The fourth hypothesis, that an anchoring effect would be observed in participants' trust on the basis of alarm reliability, was evaluated using a 3 × 12 split-plot ANOVA with the 50%, 75%, and 100% reliability level. It was predicted that there would be significantly higher mean trust scores in the 75% reliability group than the 50% reliability group. Similarly, the 100% reliability group would exhibit significantly higher mean trust scores than the 75% or 50% groups.

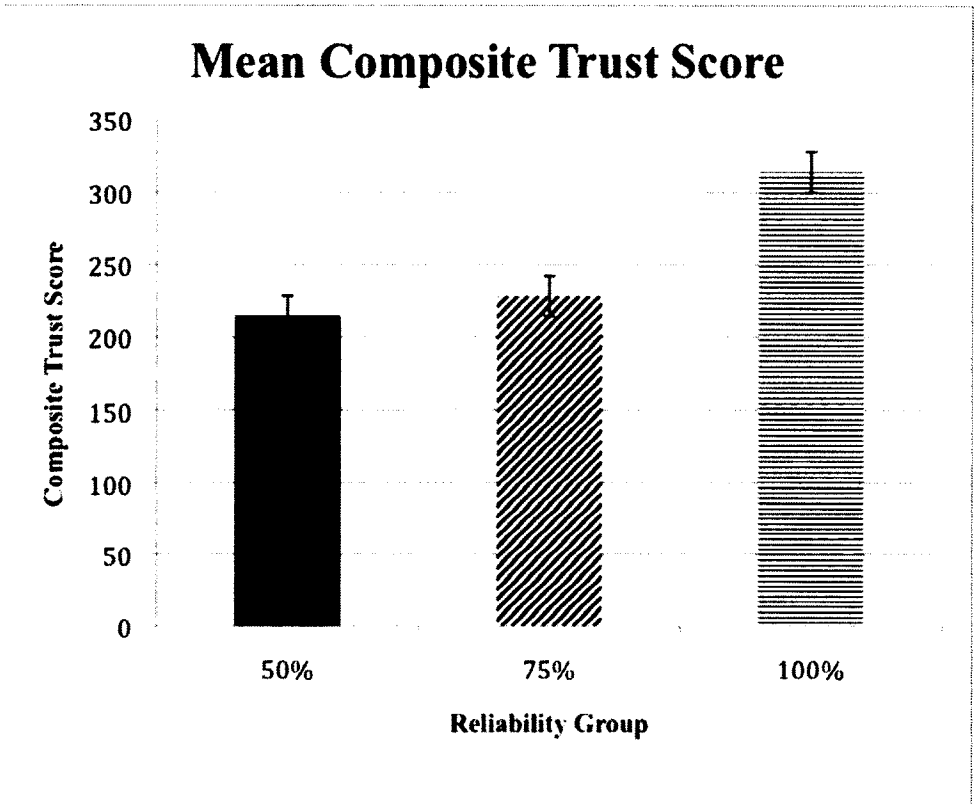

Figure 7. Mean Composite Trust Scores by Reliability Level.

There was a significant between groups difference in reliability level, $F(2, 41) = 14.60$, $p < .001$, $\eta_p^2 = .41$. Planned pairwise comparisons revealed mean trust scores were significantly higher for the 100% reliability level ($M = 314.1$, $SD = 14.3$) than for the 50% reliability level ($M = 214.9$, $SD = 13.8$), $F(1, 41) = 25.11$, $p < .001$, $\eta_p^2 = .38$. Similarly, the average 100% trust score ($M = 314.1$, $SD = 14.3$) was significantly higher than the average 75% trust score ($M = 228.6$, $SD = 13.8$), $F(1, 41) = 18.65$, $p < .001$, $\eta_p^2 = .31$. However, no significant difference in trust was found between the 50% and 75% reliability groups (see Figure 7).

CHAPTER IV

DISCUSSION

The goal of the present study was to assess the system percentage strategy identified by Gilson et al. (2001) using a larger 12-alarm array. In accordance with previous literature on reliability (Wickens & Dixon, 2007), two reliability levels were used when evaluating this strategy: 75%, and 50%. Dependent measures included reaction time, alarm acceptance rate, and scores from a multi-dimensional trust instrument. Overall, the hypotheses were supported. Subjective trust varied as a function of the overall number of additional activated alarms and was impacted by reliability level. Acceptance rate did follow a linear trend, as predicted; however, the highest order interpretable trend that was significant was cubic. Reaction time was quadratic in nature, as predicted. The implications of these results are discussed below.

Subjective versus Objective Measures of Trust

Much consideration was given to the use of subjective and objective measures of trust in this experiment. Getty et al., (1995) were among the first researchers to emphasize the use of reaction time to document trust as a function of reliability (which they termed Positive Predictive Value, "the probability that a warning will truly indicate some specified dangerous condition"; Getty et al., 1995, p. 30). Based on their study, they concluded faster reaction times are due to the higher Positive Predictive Value of an alarm. Although the definition of Positive Predictive Value can be interpreted as conditional reliability (i.e., the probability of an alarm signal given an event in the environment), the researchers also equated low Positive Predictive Value with the cry wolf effect, a phenomenon frequently associated with trust. Given this early research, it

has been assumed by some alarm researchers (e.g., Rice, 2009) that an operator will respond more quickly to an alarm they trust and consequently more slowly to an alarm they do not trust.

Interestingly, Gilson et al. (2001) found that confidence, a dimension of trust, increased linearly as a function of the number of overall activated alarms. If participants do respond more quickly to an alarm they trust, then reaction time should, therefore, also follow a linear trend. However, in the present study reaction time formed a quadratic trend. This quadratic pattern differs from the linear trend observed in the subjective trust data, indicating that reaction time may not be the best measure of trust.

These findings support the work of Wiegmann et al. (2001), who did not found any correlations between reaction time and subjective measures of trust. In his work, Wiegmann (2001) shares a similar viewpoint of Lee and See (2004) by advocating trust as a psychological construct that should be assessed only with subjective measures. Given the differing response patterns observed in the objective and subjective trust data of this study, the results of this experiment support this recommendation.

Although behavioral measures of trust represent less invasive alternatives to subjective measures, defining trust as a behavioral response should be approached with caution, as behavioral measures may reflect more than just participant trust. Moreover, the quadratic trend predicted in reaction time was based on inferences of response confidence, implying other factors may better explain the variance in response behaviors. It should be noted that response confidence is different from task self-confidence, which has be studied and found to impact automation trust (Lee and See, 2004). Task self confidence is the confidence the user has in his or her own ability to perform the

automated task. For example, a nurse with high task self-confidence in blood pressure monitoring may be less likely to use a blood-pressure monitoring automation. Response confidence, as used in this study, refers to how confident the user is in his or her response decision. These factors represent possible avenues of future research that should be explored.

Multiple Alarms on Trust and Acceptance Rate

The primary goal of this study was to examine trust and acceptance behavior when multiple alarms are present. As expected, participants calibrated their trust and acceptance rate of the test alarm based on the overall number of active alarms in the system (see Figure 5). These results are similar to those found by Gilson et al. (2001) and support a system-wide theory of trust, wherein additional alarms were found to effect response time and acceptance rate (Rice, 2009).

Notably, the analysis of acceptance behavior based on the number of additional alarms allows for greater examination of the system-wide trust theory than previously reported. The results of this study suggest that what is currently considered system-wide trust theory may simply be an extension of probability matching theory (Bliss et al., 1995).

Probability matching behavior occurs when participants match their acceptance rate to the probability of a true alarm. In multiple-alarm situations, it appears participants may employ an analogous strategy to determine alarm acceptance. The difference is the probability of alarm validity was inferred from the overall number of active alarms rather than on the given alarm system reliability. When more than 50% of the system was activated, which would indicate greater-than-chance odds, the test alarm was generally

accepted. Conversely, when the probably of a true alarm was below 50%, or less than six activated alarms, the test alarm was rejected.

Previous research with individual alarms has consistently indicated that reliability level affects trust (Lee & See, 2004; Meyer, 2001; Rice, 2009); however, the average trust ratings for the 50% and 75% reliability level in this study were not significantly different. These findings represent a departure from previous reliability research, and further suggest operators may disregard given or learned reliability information in favor of using the number of activated alarms in order to determine the probability of a true alarm.

Design Implications

The results of this study have notable implications for display design. The growth of complexity in system operations has increasingly separated the operator from the raw data used by the system, creating opaque systems in a variety of domains (Wiegmann et al., 2001). In an opaque system, information concerning raw data, system processing, and algorithms are generally unavailable to the operator. Aviation cockpits and operating rooms serve as real-world examples of these complex systems requiring multiple signaling systems. When raw data or algorithms are absent, operators are forced to evaluate a system's recommendation without understanding the basis for the recommendation. The operator must then rely on other factors, such as the probability of an alarm being true, when choosing to accept an alarm. Researchers recommend increasingly transparent designs, such as displaying the processes and algorithms involved in automation, to mitigate the detrimental effect of opaque displays on operator performance (Lee and See, 2004; Wiegmann et al., 2001). Yet, implementing such

displays may well exacerbate already high levels of workload in applied task environments.

This transparency may be even more critical for multiple alarm displays. The results of the present study suggest the operator relies heavily on the visual display to evaluate the probability of an alarm being true when multiple alarms are present. This may be a function of the mental workload and attention required when using a given reliability level or past experience for probability calibration. With a single alarm, the number of alarm presentations and the number of accurate alarms must be continuously monitored to calculate the reliability based on experience. Even if the reliability is obtained without personal experience, the operator must still monitor the overall number of alarms presented in order to sustain an acceptance percentage that approximates the probability of a true alarm. This behavior has been observed before in prior alarm research (Wiegmann et al., 2001)

The visual display of multiple alarms represents a potentially faster and less taxing alternative to calculating the probability of a true alarm, as evidenced by the use of concurrent alarm number in trust and acceptance rates for this study. Instead of monitoring a display over time, the user can make a quick estimate based only on immediately available display information. This increased dependency on the display may amplify the disadvantages opaque display design, indicating an increased need for transparency in multiple alarm displays.

The number of alarms must also be carefully considered. The attentional and temporal demands of a complex environment may limit the operator's ability to fully analyze the system recommendation, even in a transparent display. The results of this

study suggest trust in an individual alarm signal would be lower in a larger alarm array than a smaller array. Similarly, if alarms in an array are part of an unrelated subsystem processes, the operator's use of these additional alarms to calculate the probability of a true alarm may result in inappropriate levels of alarm trust and incorrect response selections.

To give a real world example, hospital rooms often contain monitors with multiple alarm displays. Consider the scenario where five alarms are co-located on a screen with a blood-oxygen alarm. The results of this study suggest that when the oxygen monitor signals, the nurse will be more less likely to consider the signal a true alarm than if only two related alarms were co-located in the display. The failure to consider the alarm a true alarm can potentially result in the nurse failing to take appropriate action. This issue is further compounded if the other co-located alarms monitor unrelated functions, because the likelihood of multiple alarms signaling is generally lower than when related functions are monitored. Thus, a given alarm is more likely to be the only alarm signaling, and therefore also more likely to be dismissed as a false alarm due to the lack of additional signaling alarms. It is recommended that in addition to transparency, the display should be limited to related and necessary alarms to alleviate the influence of simultaneous alarms on primary alarm responses.

Future Research

To more fully explore design recommendations, future researchers should evaluate the impact of specific design principles on multiple alarm arrays. The proximity compatibility principle (Wickens & Carswell, 1995) provides a guideline for display location based on perceptual processing. The stimuli used in this study were configured

based on the proximity compatibility principle: homogeneous features and the co-location of displays. Co-located displays reduce information access costs, or the costs associated with visual search and shifts in attention across a display. Similarly, homogeneous features aid in integrative processing (Wickens & Carswell, 1995). In this study, the use of homogeneous rectangles may have aided in percent calculations by allowing the participant to estimate percent activation as a function of shaded area, whereas the collocation of the alarms may have assisted in the mathematical calculation of system percentage through the reduction of visual search and information access cost.

The use of heterogeneous features to separate unrelated alarms may lessen the potential influence of simultaneous alarms on primary alarm decisions. This could be particularly valuable in complex and space-limited environments, such as airplane cockpits, where the likelihood of multiple alarms in close proximity is high. The use of visual demarcations to join related alarms may also help to separate any unconnected and unrelated alarms, thus reducing the likelihood of unrelated alarm inclusion in the primary alarm decision.

Additionally, future research should address alarm reliability levels above 75%. The reliability percentages used in this study may not have fully captured the variance of operator behavior in an unreliable system. Although there was no self-reported trust differences between the 50% and 75% reliability group, there was a difference between the 100% reliability group and the lower reliabilities respectively. These differences may suggest response behavior changes at an untested level of reliability. Using reliabilities between 75% and 100% may reveal the threshold at which reliability level is considered over the number of activated alarms present. Smoke detectors represent a real world

system in which reliability rates fall in this 75%-100% range (Bukowski, Budnick, Schemel, 1999).

Limitations

There are several limitations to this research, which should serve as additional considerations for future research in this area. First, the system used in this study is a false-alarm prone system. Considerable research has shown differing effects on human-automation interaction due to false alarm versus miss-prone systems (Parasuraman & Riley, 1997). As such, the results of this study should only be applied to similar false-alarm prone systems.

Although the trust measure used in this study enabled a multi-dimensional evaluation of trust, it has not been validated in the literature. Without validation, it is possible that another construct may be responsible for the findings. Also, it should be noted that alarm acceptance rates were calculated based on five presentations of each alarm configuration. This limits the number of possible acceptance rate values used in the data and accordingly reduces variability in responses. It is possible that with a greater number of alarm presentations, the increased variability would reveal a linear relationship, as opposed to a cubic trend.

As noted previously, the alarm stimuli used in the present study were gray in color. This coloration differs from the traditional alarm color (red). Previous research suggests that participants respond more quickly and with more force to lightness-matched red alarms than gray alarms. Consequently, faster reaction times may be elicited when using traditional alarm color. In addition to the use of red alarms, future studies should

use stimuli specific to a complex environment, such as an operating room or airplane cockpit, to increase ecological validity.

Finally, operators interact with signaling systems over long periods of time. This study does not address the possible impact of fatigue on strategy use.

Conclusions

The results of this study revealed a unique application of probability matching behavior observed with multiple alarm displays. In this experiment, the number of activated alarms was used to estimate the probability of an alarm being true, instead of a test alarm's given reliability level. This represents an important theoretical contribution through the extension of current probability matching theory (Bliss et al., 1995). Furthermore, when taken in conjunction with previous work by Gilson et al. (2001), there is evidence this strategy is employed for a variety of array sizes and alarm reliabilities. Design implications include the importance of transparent displays and limiting large alarm arrays, especially in complex and opaque environments. Specific design solutions, such as those related to the proximity compatibility principle, should be explored in future research.

REFERENCES

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

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

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.

Bukowski, R. W., Budnick, E. K., & Schemel, C. F. (1999). Estimates of the operational reliability of fire protection systems. *Proceedings of the International Conference on Fire Research and Engineering*, *3*, 87-98.

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. doi: 10.1080/1463922021000054335

Chancey, E. (2013). *The role of trust as a mediator between system characteristics and response behaviors.* (Master's Thesis). Norfolk, Virginia: Old Dominion University.

Chancey, E. T., Proaps, A., & Bliss, J. P. (2013). The role of trust as a mediator between signaling system reliability and response behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 285-289.

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,* 147-164.

Elliot, A. J., & Aarts, H. (2011). Perception of the color red enhances the force and velocity of motor output. *Emotion, 11*(2), 445-449.

Geels-Blair, K., Rice, S., & Schwark, J. (2013). Using system-wide trust theory to reveal the contagion effects of automation false alarms and misses on compliance and reliance in a simulated aviation task. *The International Journal of Aviation Psychology*, *23*(3), 245-266.

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.

Gilson, R. D., Mouloua, M., Graft, A. S., & McDonald, D. P. (2001). Behavioral influences of proximal alarms. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 595-610. doi: 10.1037/a0022599

Hollands, J.G., & Dyre, B.P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107,* 500-524.

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.

Keller, D., & Rice, S. (2009). System-wide versus component-specific trust using multiple aids. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology*, *137*(1), 114-128.

Konkai, A., Oakley, B., & Bauld, J. (2012). Reducing hospital noise: A review of medical device alarm management. *Biomedical Instrumentation and Technology*, *46*(6), 478-487. doi: 10.2345/0899-8205-46.6.478

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46* (1), 50-80.

Lynas, D., & Horberry, T. (2011). Human factor issues with automated mining equipment. *Ergonomics Open Journal*, *4*, 74-80. doi: 10.3390/min2040417

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd Edition ed.). New York, NY: Psychology Press.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

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

Moray, N. (1997). Human factors in process control. In G. Salvendy (Ed.), *Handbook of*

human factors and ergonomics, (pp. 1944–1971). New York: Wiley.

Mosier, K. L., Skitka, L. J., & Korte, K. J. (1994). Cognitive and social issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 191-197). Hillsdale, NJ: Erlbaum.

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

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429-460.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253. doi: 10.1518/001872008X288547

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

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology 49*(1), 95-112.

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

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.

Sheridan, T. B. (1981). Understanding human error and aiding human diagnostic behavior in nuclear power plants. In J. Rasmussen & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 19–35). New York: Plenum.

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

Wickens, C. D., & Carswell, C. M (1995). The proximity compatibility principle: It's psychological foundation and relevance to display design. *Human Factors, 37*, 473-494. doi: 10.1518/001872095779049408

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., Lee, J., Liu, Y., & Becker, S. G. (2004). *An introduction to human factors engineering* (2nd ed.). Upper Saddle River, NY: Prentice Hall.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, *2*(4), 352-367.

Woods, D., Sarter, N., & Billings, C. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926-1943). New York: Wiley.

APPENDIX A

INFORMED CONSENT STATEMENT

Purpose of this Form: 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.

Research Project Title: Effects of Simultaneous Alarms on Resolution Heuristics

Responsible Project Investigator(s): J. Christopher Brill, Ph.D., Assistant Professor, College of Sciences, Psychology Department

Co-Investigator(s): Amanda Allen, Graduate student, College of Sciences, Psychology Department

Overview of Research Project: This experiment is intended to examine how you judge a test alarm when other alarms are also present. If you choose to participate in this study, you will be asked to respond to the presentation of visual alarms on a computer screen.

If I choose to participate, what will I be asked to do?
You will be asked to complete a brief medical history to ensure that you are eligible to participate in the study. This medical history primarily asks about conditions or medications that might be related to sensory deficits (e.g., loss of hearing, reduced skin sensitivity) and motor ability. You may refuse to answer any questions that make you feel uncomfortable.

The researcher will then seat you in front of the computer screen, and you will be provided with more specific instructions on how to complete the task. You will have the opportunity to ask for clarification if any aspect of the task is confusing.

What steps are being taken to ensure my privacy?
All information you provide will be kept confidential, and none of the forms will list your name. This form will be separated from the rest of your data packet so no one can link your data and your identity. All written information (e.g., surveys, forms, etc.) is kept in a locked file cabinet. A numerical code will be used for all electronic information (e.g., performance data) so that your identity cannot be linked with the data file.

Are there any risks associated with participating in this experiment?
The experiment does not require you to perform actions beyond those experienced in everyday life. Therefore, this protocol is deemed minimal risk.

What if I have questions about the experiment or its procedures?
You may ask questions about the experiment at any time. If you have questions after the experiment session has ended, you may contact Dr. Chris Brill at jcbrill@odu.edu or (757) 683-4242. The ODU Institutional Review Board (ODU-IRB) has reviewed my request to conduct this project. If you have any concerns about your rights in this study, you may contact the Office of Research at (757) 683-3460 or George Maihafer of the ODU-IRB at (757) 683-4520 or email gmaihafe@odu.edu.

How long does the experiment last?

It varies from person to person, but a typical time commitment approximately 30 minutes.

Will I receive any compensation for participating in this experiment?
If you decide to participate in this study, you will receive 1 Psychology Department research credit, which may be applied to course requirements or extra credit in certain Psychology courses. Equivalent credits may be obtained in other ways. You do not have to participate in this study, or any Psychology Department study, in order to obtain this credit.

Are there any benefits or costs associated with participating in this experiment?
While there are no direct benefits for participation in this study, the results will be useful for evaluating the nature of alarm resolution. Since this study uses technology largely encountered in daily life (desktop computer, and videogame-like systems), there are no additional risks.

Is there anything else I need to know?
You must be 18 years of age or older to participate in this experiment. Additionally, in order to be eligible for participation in this study you must not have any major sensorimotor impairment that might impact your ability to perceive or respond to visual and tactile signals. You are free to withdraw from the experiment at any time without any negative consequences; however, you will only be compensated for the amount of time you spent participating in the experiment.

We will be recruiting approximately 50 participants for this study.

I have read the procedure described above. I voluntarily agree to participate in the procedure and I have received a copy of this description.

_____

Participant's Signature Date

_____

Investigator's Signature Date

APPENDIX B

DEMOGRAPHICS AND MEDICAL QUESTIONNAIRE

This survey was designed to obtain information about our research participants prior to serving in our studies. We need this information to help us interpret your results. ALL data collected in this laboratory is to be kept confidential.

1) Age: _____

2) Sex (circle one): Male / Female

3) Handedness: Left / Right

4) Do you have any medical conditions or injuries affecting your vision? Yes / No

   4a) If yes, please explain:

   _____

   4b) If applicable, did you bring a correction with you? (i.e., glasses or contact lenses): Yes / No

5) Do you have any medical conditions or injuries affecting your hearing? Yes / No

   5a) If yes, please explain:

   _____

6) Do you have any medical conditions or injuries affecting your sensitivity to touch? Yes / No

   6a) If yes, please explain:

   _____

7) Do you have any medical conditions or injuries affecting your motor control, particularly the use of your hands? Yes / No

   7a) If yes, please explain:

   _____

8) Do you have any medical conditions affecting your ability to pay attention? Yes / No

8a) If yes, please explain:

_____

9) How often do you play video/computer games? Never Monthly Weekly Daily

9a) If you do play video/computer games, circle the number that corresponds to how **confident** you are using video/computer **games**:

1	2	3	4	5	6	7
Low			Average			High

APPENDIX C

DESCRIPTIVE STATISTICS FOR EXPERIMENTAL DATA

Descriptive Statistics for Acceptance Rate by Alarm Level for the 50% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	0.00	1.00	0.19	0.08	2.22	5.98
2	0.00	1.00	0.21	0.10	1.48	1.06
3	0.00	1.00	0.27	0.08	1.45	1.45
4	0.00	1.00	0.33	0.09	0.78	-0.85
5	0.00	1.00	0.37	0.10	0.65	-1.04
6	0.00	1.00	0.55	0.08	-0.08	-0.95
7	0.20	1.00	0.77	0.08	-1.06	0.14
8	0.40	1.00	0.85	0.06	-1.35	0.59
9	0.00	1.00	0.81	0.08	-1.88	2.76
10	0.20	1.00	0.87	0.07	-1.98	3.82
11	0.40	1.00	0.92	0.06	-2.31	4.66
12	0.60	1.00	0.89	0.07	-1.16	-0.41

Descriptive Statistics for Acceptance Rate by Alarm Level for the 75% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	0.00	1.00	0.24	0.08	1.15	-0.06
2	0.00	1.00	0.35	0.10	0.60	-1.45
3	0.00	0.80	0.20	0.08	1.10	-0.47
4	0.00	1.00	0.33	0.09	0.64	-0.35
5	0.00	1.00	0.53	0.10	-0.28	-1.44
6	0.40	1.00	0.82	0.08	-0.83	-1.14
7	0.00	1.00	0.76	0.08	-1.27	0.58
8	0.40	1.00	0.80	0.06	-0.75	-1.19
9	0.20	1.00	0.81	0.08	-1.49	0.78
10	0.20	1.00	0.79	0.07	-1.20	0.37
11	0.20	1.00	0.86	0.06	-1.74	1.78
12	0.00	1.00	0.85	0.07	-2.11	3.36

Descriptive Statistics for Acceptance Rate by Alarm Level for the 100% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skew*	*Kurtosis*
1	0.00	1.00	0.70	0.38	-1.08	-0.41
2	0.00	1.00	0.73	0.34	-1.01	-0.01
3	0.00	1.00	0.73	0.37	-1.13	0.08
4	0.00	1.00	0.77	0.32	-1.54	1.64
5	0.00	1.00	0.79	0.34	-1.56	1.26
6	0.20	1.00	0.89	0.22	-2.70	8.26
7	0.40	1.00	0.87	0.22	-1.70	1.82
8	0.40	1.00	0.90	0.19	-1.94	3.18
9	0.20	1.00	0.89	0.24	-2.17	4.25
10	0.60	1.00	0.91	0.15	-1.53	0.94
11	0.00	1.00	0.86	0.28	-2.64	7.62
12	0.60	1.00	0.93	0.15	-1.87	2.09

Descriptive Statistics for Reaction Time by Alarm Level for the 50% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	1244.80	3828.00	2020.34	669.99	1.55	3.22
2	1305.80	4003.60	2046.99	800.95	1.29	1.24
3	1086.00	6013.20	2624.94	1341.53	1.57	2.18
4	1172.40	31135.80	5930.07	7715.28	3.04	10.18
5	1129.20	7906.00	3575.67	2206.27	0.94	0.07
6	1366.20	11115.00	4405.97	2634.15	1.27	2.18
7	1096.00	8365.00	4264.56	2135.56	0.32	-0.43
8	1464.40	16690.40	4184.71	4032.71	2.67	7.72
9	1294.00	14576.60	3659.93	3654.45	2.55	6.46
10	1239.40	7592.20	2590.19	1562.02	2.83	9.17
11	1067.80	3732.20	2344.50	849.24	0.26	-1.12
12	1227.40	3855.60	2154.71	801.35	0.86	-0.05

Descriptive Statistics for Reaction Time by Alarm Level for the 75% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	1171.60	11697.60	2929.79	2896.55	2.49	6.17
2	1293.40	7047.20	2716.04	1469.46	2.10	5.08
3	1107.00	16289.40	3918.92	4463.91	2.08	3.68
4	1524.00	12126.40	3501.39	2783.24	2.48	6.65
5	1400.20	8562.00	3207.29	1970.21	1.75	3.09
6	1318.00	16508.80	4275.83	4042.12	2.33	5.85
7	1236.60	10468.60	3543.28	2410.97	1.84	4.19
8	1358.20	10065.40	3311.97	2839.74	1.76	1.93
9	1403.60	9538.80	3430.29	2802.50	1.47	0.77
10	1038.60	8164.20	2641.24	1931.84	1.99	4.16
11	1078.20	6561.00	2409.61	1436.67	1.89	4.24
12	1237.80	6677.20	2254.92	1375.60	2.62	8.14

Descriptive Statistics for Reaction Time by Alarm Level for the 100% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	1023.80	4631.00	2866.84	1212.41	0.26	-1.14
2	649.00	4813.80	2639.44	1206.26	0.17	-0.52
3	741.20	7878.80	3504.39	2149.44	0.66	0.05
4	682.60	8570.40	3157.97	2205.70	1.54	2.19
5	846.40	6265.80	3269.29	1485.79	0.23	-0.21
6	746.40	8810.40	3090.21	1945.01	2.02	5.79
7	1090.60	12104.40	3853.86	2953.19	1.77	4.00
8	958.20	6972.80	3302.91	2258.30	0.60	-1.57
9	1137.40	6681.80	3193.29	1645.10	0.94	0.14
10	1064.60	14819.00	3858.07	3627.78	2.41	6.57
11	905.60	9923.80	3322.89	2497.44	1.56	2.72
12	933.60	4084.20	2217.33	928.01	0.53	-0.49

Descriptive Statistics for Composite Trust by Alarm Level for the 50% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	10.00	229.60	134.33	62.24	-0.40	-0.16
2	12.00	307.00	165.44	70.58	-0.22	0.90
3	85.50	282.00	171.47	48.72	0.29	0.99
4	89.25	236.00	178.67	35.66	-0.90	1.75
5	136.00	231.00	188.80	26.27	-0.55	0.46
6	144.00	248.70	205.51	27.29	-0.86	1.18
7	131.00	240.00	213.56	25.61	-2.58	8.35
8	200.00	285.00	241.59	27.38	-0.14	-1.06
9	184.00	314.00	240.95	36.78	0.19	-0.55
10	193.00	451.20	267.53	68.54	1.39	2.43
11	200.00	375.00	286.68	57.92	0.14	-1.37
12	128.20	400.00	283.83	80.29	0.01	-0.50

Descriptive Statistics for Composite Trust by Alarm Level for the 75% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
1	14.00	296.00	128.77	92.11	0.33	-1.26
2	37.80	309.80	176.99	93.11	-0.13	-1.44
3	53.60	290.00	170.15	66.28	-0.12	-0.70
4	117.00	276.00	192.56	42.16	0.29	-0.26
5	162.00	272.00	220.13	33.37	-0.38	-0.84
6	167.80	280.00	234.62	31.86	-0.69	-0.19
7	155.40	315.20	244.36	41.62	-0.55	0.37
8	146.75	430.40	265.52	66.07	0.64	2.03
9	82.80	327.00	256.32	67.16	-1.37	1.96
10	116.50	360.00	267.99	64.29	-0.92	0.77
11	121.00	380.00	285.92	81.67	-0.92	-0.09
12	87.00	400.00	299.59	95.84	-0.98	-0.12

Descriptive Statistics for Composite Trust by Alarm Level for the 100% Reliability Group

Number of Alarms	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*	*Skew*	*Kurtosis*
1	3.00	400.00	266.24	152.04	-0.79	-1.11
2	19.75	400.00	280.56	136.96	-1.08	-0.20
3	60.00	400.00	284.65	121.80	-0.71	-0.81
4	87.00	400.00	308.22	116.40	-1.02	-0.63
5	132.00	400.00	295.84	109.90	-0.51	-1.71
6	144.00	400.00	313.52	96.05	-0.60	-1.36
7	131.00	400.00	313.42	93.34	-0.81	-0.69
8	151.40	400.00	329.60	83.51	-0.98	-0.19
9	184.00	400.00	334.86	76.08	-0.88	-0.21
10	136.40	400.00	335.79	81.77	-1.53	1.89
11	203.85	439.00	358.72	70.12	-1.45	1.27
12	128.20	400.00	348.03	89.92	-2.00	2.97

VITA

Amanda C. Allen
1054 Anna Knapp Blvd 5B
Mount Pleasant, SC 29464
Phone: 843-902-3932
aalle044@odu.edu

Education

Old Dominion University, Norfolk, VA
2014 M.S. in Experimental Psychology, Human Factors program, *In Progress*

Clemson University
2010 B.A. in Psychology, East Asian Studies minor

Research Experience

Old Dominion University, Norfolk, VA
2012-present Applied Sensory Psychology Laboratory Research Assistant
P.I.: Dr. J. Christopher Brill

A Multi-Modal Assessment of Reserve Attention Capacity in Distracted Drivers

- Development of experimental protocols using Superlab and Tactor SDK
- Preparation of IRB applications

Work Experience

SA Technologies
2010-2011 Research Associate

- Redesign large screen concepts to work on the screen of a small handheld device
- Oversee both formal and informal reviews of usability test results to help determine the results' impact on interface design
- Evaluate and edit cognitive walkthrough and usability test plans
- Create and edit the Usability Test Plan and Report templates for other researchers to use
- Develop an internal system to track and link usability test findings to all applicable UI designs

Presentations

Allen, A., & Brill, J. C. (2013). *A Multi-modal Assessment of Reserve Attention Capacity in Distracted Drivers.* Paper presented at the annual meeting of the Human Factors and Ergonomics Society, San Diego, CA.