This is to certify that the
thesis entitled

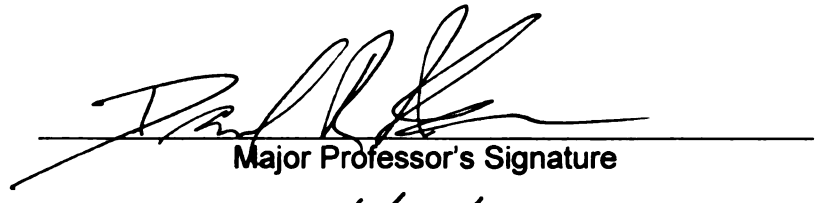
THE EFFECTS OF PERSON AND MACHINE CHARACTERISTICS ON OPERATOR USE AND MONITORING OF AN AUTOMATED SYSTEM

presented by

Stephanie M. Drzakowski

has been accepted towards fulfillment
of the requirements for the

M.A. degree in Psychology



Major Professor's Signature

4/26/05

Date

THE EFFECTS OF PERSON AND MACHINE CHARACTERISTICS ON OPERATOR USE AND MONITORING OF AN AUTOMATED SYSTEM

By

Stephanie M. Drzakowski

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
For the degree of

MASTER OF ARTS

Department of Psychology

2005

ABSTRACT

THE EFFECTS OF PERSON AND MACHINE CHARACTERISTICS ON OPERATOR USE AND MONITORING OF AN AUTOMATED SYSTEM

By

Stephanie M. Drzakowski

Organizations are increasingly automating tasks that were once performed solely by humans. However, in many cases automation has failed to increase organizational productivity. This study posits that human operators of automated technology make errors at two decision points: the decision of whether to use the automated system and the decision of whether to monitor the automated system for errors. The present research investigates the factors influencing these two decisions. A sample of 253 undergraduates participated in a computer-based simulation of an airport security screening task. While visually searching for guns and knives in X-ray images of luggage, participants had the option of using an automatic screener. Results indicated positive and significant effects of trust and affect toward the automated system on use and monitoring of this automatic screener. The effects of self-efficacy, conscientiousness, extraversion, desire to decrease cognitive load, and the interaction between trust and self-efficacy were non-significant.

ACKNOWLEDGMENTS

First and foremost, Dan, this project never would have happened without your tremendous support. Thank you for encouraging me to pursue this topic despite the fact that I was originally unsure about its feasibility and place in the I/O literature. I'm so glad that you convinced me to choose a topic I was truly interested in. In addition, I am incredibly grateful for your assistance in developing the task simulation. Thank you for giving so willingly of your resources. I hope that the program will continue to be useful for you and others in the future. And most of all, thank you for the enormous amount of time and effort you dedicated to reading and providing comments on my numerous drafts! Your helpful feedback not only infinitely improved the project, but also helped me grow as a researcher. You have been an excellent model of careful thought and clear articulation of ideas. I have learned so much from working with you.

Neal and Fred, thank you so much for agreeing to serve on my committee. You were there when I had questions or concerns, and your insightful feedback on the project has been tremendously helpful. Thank you so much again.

I'd also like to thank the Transportation Security Administration. The TSA provided X-ray images without which the experimental task used in this study would not have been possible. Much thanks to the TSA for its help and support.

Dan N., thank you so much for helping me out with the slides. All of those hours you spent cutting and pasting X-ray images were much appreciated! Mike V., I'd also like to thank you for your help in testing the program, running sessions, and checking all

of the data in that Excel file. Thanks to both of you for doing such a great job – good luck with your future endeavors!

Mom and Dad, thanks for supporting me through the rough times – especially through the first year I was here. Your encouragement meant so much, and now look—your daughter has finally finished her master's thesis! Thank you so much for listening to me whine and for always supporting me through all of this process.

Also, thanks to Alyssa, Smriti, and Dan W. for volunteering your time to help me test my program. Your input was much appreciated! Anna, thanks for all those great TV / thesis nights last year...and thanks so much for lending an ear and making me laugh when I needed it. And a big thanks to all of the grad students – it was great to have the support of such a great group!

Finally, I'd like to thank my husband. Kevin, thanks for being there for me through the good times and the bad. Thanks for comforting me when I was upset and celebrating with me when I reached a milestone in the process (who could forget the Thesis Peep?). Most of all, thank you for agreeing to move to Michigan for three years! That's quite a sacrifice, and I hope you know how much I appreciate it. But now the race is on! You may have beaten me by a year on the M.A., but I've finally caught up to your educational level. Who will finish the Ph.D. first? Only time will tell!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

The past half-century has borne witness to an unprecedented explosion of autonomous technology development and implementation, particularly in workplace settings. Many jobs once performed by humans are now being automated. Automation is defined as "the allocation of functions to machines that would otherwise be allocated to humans, or the machines that perform those functions" (Funk, et al., 1999). For example, people once made bank deposits and withdrew cash with the aid of a human bank teller. Now, the tasks of withdrawing and depositing money are often conducted by automatic teller machines (ATMs). Telephone operators have, in many organizations, been replaced by automated directories. Physically dangerous assembly work, once undertaken at great risk, is now regularly performed by robotic systems.

Expert systems, a type of automated decision aid, have been implemented in many work environments. Expert systems are "computer programs that apply substantial knowledge of specific areas of expertise to the problem solving process" (Bobrow, Mittal, & Stefik, 1986, pp. 880). These systems are provided an expansive knowledge base in a very specific area. One of the most famous expert systems is Deep Blue (IBM), the expert system that defeated chess grandmaster Gary Kasperov (IBM, *n.d.*). Deep Blue is a highly-developed prototype system with power surpassing most machines. However, Waterman (1986) states that less powerful automated decision aid systems have been implemented in almost every type of workplace (as cited in Will, 1991). For example, they have been widely implemented in the field of behavioral accounting, where they produce recommendations for whether applicants should receive loan renewals. In addition, expert systems have been used since the 1970s in medical

diagnoses, where doctors can enter symptom information and the system produces a likely diagnosis (Yu, et al., 1979).

As more tasks have become automated, a unique situation has arisen: the situation where humans and machines must work jointly on the same task. For example, airplane cockpits are now partially automated, such that the craft have autopilot and automonitoring systems. However, we still have human pilots and co-pilots on every flight. Often, both the human pilots and autopilot system perform the same task – piloting the plane. Control can be allocated to either human or machine, but both must function correctly and work together to make flights safer and smoother. This type of joint human-machine responsibility for performance has become increasingly common (Parasuraman & Riley, 1997) and requires research attention.

Organizations spend a large amount of revenue purchasing and maintaining automated systems and training employees to use them. In return for their investments, they expect the benefits that automated systems promise. These systems are designed to reduce labor costs and to free humans from mundane and/or dangerous tasks (i.e., by reducing fatigue and boredom and by removing humans from dangerous working conditions). In many cases, they are effective. Research has established that in many decision situations, even relatively simple algorithms produce consistently more accurate results than do human judgments (Dawes, Faust, & Meehl, 1989; Leli & Filskov, 1984).

However, sometimes the implementation of automated technology does not tender the degree of benefit for which it was designed. The importance of empirical attention to human-machine interaction is highlighted by the "productivity paradox" (e.g., Lee & Perry, 1999; Stratopoulos & Dehning, 2000). The productivity paradox is that increased

implementation of advanced technology in the workplace has not consistently been significantly correlated with increased organizational performance. In fact, research has suggested that in some industries, disappointment with the degree of improvement after automation implementation tends to be common (Arnold & Sutton, 1998). It has been suggested that the disappointing degrees of improvement are not due to mechanical malfunctions or to poor design. Instead, it has been proposed that automation implementation failures are due to faulty human-automation interactions (e.g., Muir, 1994; de Vries, Midden, & Bouwhuis, 2003; Will, 1991).

Numerous fatal disasters have implicated poor human-automation interaction as a cause. For example, Sparaco (1994) describes a fatal airplane crash that could have been prevented had the pilots taken action to override the autopilot only four seconds sooner. However, overconfident in the plane's automated technology, they were too late to realize the error. Poor human-computer interaction has also been implicated in medical patient overdoses (Neuman, 1986; Sollins, 1986) and the nuclear incident at Three Mile Island (Connors, Harrison, & Summit, 1994).

In order for automated technology to be beneficial, users of the automated systems must make the correct choices about how to interact with the technology. It is argued here that there are two points at which a decision regarding the technology must be made. An incorrect decision at either of these points can severely decrease total system effectiveness.

The first decision point for the human is whether or not to *use* an automated system. The major error at this decision point is that the user may fail to use the automated system when doing so would improve performance. For example, a doctor

may fail to use an expert system for advice on a diagnosis even when the system has a much higher likelihood of making a successful diagnosis than does the doctor.

This second decision point is only encountered after one chooses to use the automated system. At this point, users are required to determine what actions they will take after engaging the system. They decide whether to *monitor* the system for errors so that they can override the system if an error occurs. Automated systems, like humans, sometimes make mistakes or non-optimal decisions. Often the algorithms on which decision aids are based are probabilistic. Thus, part of the role of the human in shared decision situations is to monitor and evaluate the performance of the automated system.

The purpose of the present study was to examine the factors that might affect the decisions of humans working in conjunction with an automated system at these two decision points. The vehicle for doing so was the task of airport security screening. A simulation of the baggage screening task was produced in which participants encountered the two decision points: they were able to choose whether or not to use an "automated weapons detector," for assistance, and if they did, they had to choose whether to monitor its decisions for errors.

Literature Review

Many outcomes of interest in psychology are affected by two classes of variables, person variables and situation variables. The present study focuses on building an integrative model of some person factors that may influence automation use and monitoring. Previous research has established that person factors such as trust toward the automation, self-efficacy, and personality may be important determinants of operator choices when using automation (e.g., Dzindolet, et al., 2001; Dzindolet, et al., 2003; Lee

& Moray, 1994; Parasuraman & Byrne, 2003). The person factors relevant to the decisions operators make at the two decision points of use and monitoring fall into two categories: attitudes toward the automated system and dispositions. Attitudes toward the automated system are heavily influenced by the characteristics of the automated system, while dispositions are more stable propensities that exert effects independent of the machine's characteristics. Each of these categories will be discussed in turn.

Attitudes Toward the Automated System

Trust. The first factor upon which I focus is trust in the automated system. If the operator does not trust a system, he or she will probably resist using it. In regard to monitoring, excessive trust in the system might cause the operator to become less vigilant or reluctant to monitor or override the system. Will (1991) proposes that an implicit trust forms when a user decides to act upon the suggestions of an automated system. Furthermore, Muir (1987) states that when operators are forced to use a system they do not trust, they often go to extremes to avoid using the system; however, when users trust automation to an excessive level, they may allocate tasks to automated systems inappropriately. In order to understand the role of trust in human-computer performance, one must explore the nature of trust. I begin with an examination of theories of trust between humans. Then, a theory of trust between a human and an automated system is described. Finally, the conceptualization of trust used in the present study will be presented.

Definitions of Trust. Some studies that examine trust do not provide a definition of the construct but instead leave the reader to utilize his or her own personal and colloquial definition of the term (e.g., Dzindolet, et al., 2003; Lee & Moray, 1994;

Lewandowski, Mundy, & Tan, 2000). Among those researchers who do provide explicit definitions of trust, little consensus is found. Table 1 presents a few of the numerous disparate definitions of trust that can be found in the literature.

Table 1

Selected Definitions of Trust

Source	Definition
Barber (1983)	"socially learned and socially confirmed expectations that people have of each other, of the organizations and institutions in which they live, and of the natural and moral social orders that set the fundamental understandings for their lives"
Hosmer (1995)	"the expectation...of ethically justifiable behavior – that is, morally correct decisions and actions based upon ethical principles of analysis"
Lewis & Weigert (1985)	"undertaking of a risky course of action on the confident expectation that all persons involved in the action will act competently and dutifully"
Robinson (1996)	"expectations, assumptions, or beliefs about the likelihood that another's future actions will be beneficial, favorable, or at least not detrimental to one's interests"

Not all of these definitions appear to be relevant to the study of human-computer trust. In order to develop an appropriate definition of trust in the context of the current study, the literature on social trust between humans and human-computer trust was reviewed. This review resulted in the identification of the components of trust that appear most applicable to human-machine trust relationships.

Researchers studying human-computer trust often apply theories of social trust between humans to their research (e.g., Lee & Moray, 1992; Muir, 1987; Muir & Moray, 1996). Evidence suggests that elements of social trust and human-computer trust are often similar and thus, research and theory in one area can reasonably be applied to research in the other. Jian, Bisantz, and Drury (2000) specifically suggest that findings on interpersonal trust are applicable to human-machine trust. Their study asked participants to perform a sorting task in which they identified and ranked words as being associated with trust between people, trust between and human and an automated system, and trust in general. They compared the sets of the top 15 words that participants judged to be related to trust between two humans and trust between a human and a machine. They found the union set to be small – seven words (whereas five words would indicate absolute agreement). This suggests that people perceive human-human trust and human-machine trust as being very similar. Additionally, of the top five words in each set, three (trustworthy, loyalty, and reliability) were identical (see Table 2).

One theory of trust that is directly focused on automated systems has been proposed (Muir, 1987). This theory was based heavily on two of the most popular theories of social trust between humans: Rempel, Holmes, and Zanna's (1985) theory and Barber's (1983) theory. In the following paragraphs, I discuss these two theories in depth. Following the discussion of the theories, I turn to Muir's theory of trust in expert systems and the research supporting it.

Rempel, Holmes', & Zanna's Theory. Perhaps the most popular theory of trust was proposed by Rempel, Holmes, and Zanna (1985). This theory, which was developed in order to describe close interpersonal relationships, does not suggest one specific

definition of trust but instead conceptualizes trust as a hierarchy of attributions one may make about one's partner. In this hierarchy, each level becomes more abstract than the last. The first level, predictability, concerns specific patterns of behavior. When one's partner is highly predictable, one can forecast how one's partner will react in a given set of circumstances. At the predictability stage, little effort is made to determine *why* any specific behavior is exhibited; predictability simply entails a consistent response in a given context.

Table 2

Top Five Words for Trust in General, Trust Between Humans, and Trust in a Machine

Trust in General	Trust Between Humans	Trust in a Machine
Trustworthy	Trustworthy	Trustworthy
Honesty	Honesty	
Loyalty	Loyalty	Loyalty
Reliability	Reliability	Reliability
Honor		Honor
	Integrity	
		Familiarity

In the second stage of the trust hierarchy, dependability, one makes attributions about the partner's stable disposition. The focus is no longer on specific behaviors in specific situations; the focus rests on the partner him- or herself. One may say, "My partner is a dependable person. I can trust him or her not to hurt me." Dependability is distinct from predictability in that it is a general attribution about the referent person, not an attribution about a context-specific behavior.

Faith, the third and final stage in the trust hierarchy, is the most abstract. It involves confidence that, in the face of uncertain future events, the partner will behave in appropriate ways. It can be considered a state of emotional security where one believes that his or her partner's motives and intentions are righteous, such that in any situation the partner will behave appropriately. Like dependability, faith is an attribution about the person. However, unlike dependability, faith extends into the future and to situations that are completely novel; thus, it is not necessarily based on any specific evidence from past experience.

It should be noted that Rempel, et al.'s theory specifies that trust develops over time. As a relationship matures, trust usually evolves from predictability to dependability to faith. According to Rempel, et al., it is rare to find trust at the faith level in the early stages of a relationship. Additionally, while the stages are distinct, they are not independent. Although predictable behavior is not sufficient for the development of dependability, it serves as evidence of a dependable disposition. Therefore, it is unlikely that someone considered unpredictable would be considered dependable, just as it is unlikely that someone considered undependable would be trusted at the faith level.

In summary, Rempel, et al.'s theory suggests that trust develops over time and progresses through stages of increasing abstraction, from predictability of specific behaviors to attributions of a dependable character to faith in future behavior regardless of context. Because they are based on actions occurring in known contexts, the concepts of predictability and dependability can easily be applied to the concept of trust in automation. An automated system can function consistently in a known situation – so much so that the user might attribute dependability to the machine. The concept of faith,

however, is more difficult to conceptualize as related to machines because it implies that the user will generalize the machine's dependability to new, future situations.

Barber's Theory. In 1983, Barber proposed his influential theory of trust. This theory, which is sociological in nature, posits that trust exists in order to provide social ordering and control. He views trust as a function of three types of expectations that social actors have of one another as they make choices about which behaviors are rationally effective and morally appropriate. He suggests that trust develops as a function of three expectations.

The first expectation is that the natural and moral orders will persist. In this, Barber means that we expect order, regularity, and stability to some degree. We expect that social norms will remain relatively unchanged from day to day, as will the laws of physics. These expectations, according to the theory, prevent us from being paralyzed by fear of every imaginable threat, and they allow us to develop trust in others.

Secondly, trust is based on the expectation of technically competent role performance. In order for a person to trust a target other, the target must possess sufficient knowledge and technical facility to meet performance standards. In other words, the expectation reflects the simple notion of consistently producing the correct answer or product. For example, a technically competent expert system would produce the correct answer to the problem time after time.

Finally, trust is based on the expectation that target others will fulfill their fiduciary responsibility, or the responsibility to place others' interests above their own personal gain. This expectation is inherent in situations where one partner makes use of a special knowledge or skill that other members of the social system do not possess.

According to Barber's theory, this type of trust functions to prevent abuse of power, and he primarily applies the concept to politicians and doctors.

Trust in Automation: Muir's Theory of Trust in Expert Systems.

Muir (1987) provides a theory designed specifically to describe human trust in an expert system. Because the present study employs a simulation of an expert system task, Muir's (1987) theory appears particularly relevant. Muir integrates both Barber (1983) and Rempel et al. (1985) to create her theory of human trust in expert systems.

As discussed previously, expert systems are automated decision aids that produce recommendations for action in an area of expertise. Muir uses Barber's facets of trust as a base point. She then applies Rempel et al.'s concepts of predictability, dependability, and faith to discuss how trust may evolve over time in a human-computer relationship. The synthesis of Barber and Rempel et al.'s theories results in a three by three matrix into which various types of trust can be mapped (see Table 3).

In her theory, Muir also discusses how each of the six theory components, which were developed with social trust in mind, can be applied to human-computer trust relationships.

Muir begins with Barber's theory. The first component in Barber's theory is persistence, the expectation that the existing natural and social orders will persist into the future. This persistence provides some sense of environmental constancy, allowing trust to develop. Without the expectation of persistence, we would not be able to assume that the laws of gravity will persist day after day, that the word "hello" will remain a greeting tomorrow and thereafter, or that someone who we trust today will be trustworthy tomorrow. This expectation of persistence is assumed to be present in all individuals, but

the degree to which individuals expect persistence is thought to vary. Some authors have conceptualized persistence as an individual difference characteristic reflecting a disposition to trust machines (Muir & Moray, 1996). This conceptualization is adopted here and will be discussed further in the following sections.

Table 3

Basis of Trust Expectations at Different Levels of Experience

	Basis of expectation at different levels of experience		
Expectation	Predictability (of acts)	Dependability (of dispositions)	Faith (in motives)
Persistence			
Natural physical	Events conform to natural laws	Nature is lawful	Natural laws are consistent
Natural biological	Human life has survived	Human survival is lawful	Human life will survive
Moral social	Humans and computer act 'decently'	Humans and computers are 'good' and 'decent' by nature	Human and computers will continue to be 'good' and 'decent' in the future
Technical competence	X's behavior is predictable	X has a dependable nature	X will continue to be dependable in the future
Fiduciary responsibility	X's behavior is consistently responsible	X has a responsible nature	X will continue to be responsible in the future

Barber's second component, competence, refers to the ability to meet or exceed performance standards. This same concept can be easily extended to include automated systems. Automation that is perceived as competent will consistently produce desirable performance results (Lee & Moray, 1992).

Fiduciary responsibility, Barber's final component, refers to one's obligations to place the well-being of others above one's own personal gain. The opportunity to do so occurs when one member of a group possesses knowledge, skills, or expertise exceeding that possessed by other group members (Barber, 1983). Barber discusses trust at a macro level, and his conceptualization of responsibility was termed "fiduciary" because it was

framed in terms of entire professions, such as politicians' responsibility to place the welfare of their constituents above their own personal welfare. Today, although human operators sometimes act based almost entirely upon the recommendations of machines, the ultimate responsibility for those actions remains with the human operator. Machines are not blamed for accidents or for financial loss resulting from their use; instead, the blame is attributed to the operator. Therefore, the concept of fiduciary responsibility does not fit machines at the present time, and the conceptualization of responsibility must be slightly amended. Muir amends the meaning of "fiduciary responsibility" to avoid financial connotations, but she retains the adjective "fiduciary."

In order to adapt fiduciary responsibility to the human-computer context, Muir asserts that because expert systems are usually designed to provide the expert knowledge not possessed by the human operator, the human typically has little ability to evaluate whether the expert system has produced the correct recommendation. This situation sets the stage for the influence of responsibility. Responsibility has been conceptualized two ways in the human-computer trust literature thus far. First, responsibility has been discussed in terms of the programmer's responsibility to provide the best possible product (Muir & Moray, 1996). According to this conceptualization, a highly responsible machine would be impeccably designed and programmed, whereas an irresponsible machine would be designed with a less effective program. Second, responsibility has been conceptualized as the extent to which the machine gives clear explanations of how it is working and why it is requesting particular information (Muir, 1987). This second conceptualization reflects responsibility in that the machine reduces the knowledge gap between itself and the operator by making its processes and intentions explicit.

Muir next moves to Rempel, Holmes, and Zanna's (1985) components of trust, the first of which is predictability. Muir sees predictability as both a construct in itself and also as a stage of trust. As a construct, predictability receives little modification from Muir – it refers to the extent to which the machine behaves consistently in a given situation. In the predictability phase, trust is based on the consistency of one's behavior. In the case of human-computer interaction, trust will be based on the consistency of the machine's output.

At the dependability phase, the focus is on an attribution about the machine's character. In order for a machine to be considered dependable, it needs to have exhibited predictable behavior over time. For a machine, a predictable response might be to process input and produce output. When the machine converts input to output over time, the machine could be considered to be dependable. Any situation in which the machine does not consistently perform (e.g., the machine "freezes up" or produces error messages) may lead to perceptions that the machine is not dependable.

Muir also finds a way to apply faith to the human-machine relationship. In the faith stage, the emphasis is on the operator's expectation that the machine will continue to be predictable and dependable in the future. In this phase of social trust, Muir proposes that weight is given to events that indicate the person's intrinsic motivation to maintain the positive social relationship. While Muir balks at the attribution of motivation to today's machines, she does propose that human operators have faith in machines. She believes that the fact that people continually use expert systems despite an awareness that the systems operate largely beyond their understanding and can easily malfunction indicates a sense of faith in those systems. Faith, then, can be

conceptualized as the extent to which operators are *in*sensitive to cues indicating that the machine is unpredictable or undependable.

Muir's theory has received some empirical support. Muir and Moray (1996) tested the theory in two studies that simulated the operation of a milk pasteurization plant. In Study 1, participants were responsible for the supervision of three pump functions, one of which they were able to toggle between manual and automatic control.

As previously discussed, persistence is seen as an individual difference characteristic in human-computer trust relationships. Muir and Moray (1996) excluded persistence from their study due to their desire to examine situational effects only. Of Barber's two remaining trust factors, competence emerged as the greatest predictor of subjective trust, accounting for 81% of the variance in trust. Adding responsibility to the regression equation did not result in a significant improvement, although this may have been the result of a measurement artifact; the items assessing competence and responsibility were phrased almost identically. Therefore, the conclusion that competence best reflects how people define "trust in a machine" requires replication by future researchers.

Because participants had strongly preferred manual control over automatic control in Study 1, Study 2 imposed a small penalty for the use of manual control. This penalty was designed to be large enough to encourage the participants to select automatic control when the automation was functioning correctly, but small enough so that they would not hesitate to override faulty automation. Results indicated a positive correlation between subjective trust in the automation and time spent in automatic mode ($r = .71$). This

finding supports the assertion that operators will be more willing to use automation that they trust.

Furthermore, the results showed that the more error-prone the automation, the more the operators will monitor it. Operators were significantly more likely to check the status of the automatic pump when that pump had a higher fault rate. The results of this study therefore supported predictions made by Sheridan and Hennessy (1984) that an inverse relationship would exist between trust and monitoring of automation.

In summary, Muir and Moray (1996)'s study found that high trust in the automated system increases system use and decreases system monitoring. It was also found that automation competence is the factor that most influences human-machine trust, while responsibility did not significantly add to the prediction of trust.

Lee and Moray (1992) conducted a study designed to replicate and extend Muir and Moray's (1996) findings. They used a nearly identical simulation of a juice pasteurization plant. The difference was that in this study, operators could allocate all three sub-systems to any combination of manual or automatic control. Surprisingly, the results of the study failed to support predictions by Muir (1987) and Zuboff (1988) that trust will be positively associated with increased use of the automated system. Instead, they found the opposite relationship; as trust *decreased*, use of the automatic function *increased.* The authors conclude that trust is not solely responsible for influencing the allocation of automatic and manual functions. They speculate that because many participants were operating under manual control when errors were introduced into the system, the errors might have affected the allocation strategy the operator was using at the time. If the operator was using manual control of the pump when the errors occurred

(as the majority were), the operator might have been motivated to explore an alternative strategy – use of the automation. Lee and Moray suggest that automation use might be governed by a pairing of trust in the automation and operator task self-efficacy. This proposition will be discussed further later in this paper.

The present study will attempt to further illuminate the relationships between trust and automation use and between trust and automation monitoring. Lee and Moray (1992) suggest that moderators such as self-efficacy exist in the relationship between trust and use; however, the possibility that moderator variables might exist in the trust / monitoring relationship has yet to be examined.

Trust summary. In this study, trust will be conceptualized based on the suggestions presented by Muir (1987). Therefore, in our study, the components of trust will be conceptualized as follows:

Persistence: The belief that the natural and moral orders will persist, as defined by Barber (1983), cannot be manipulated for study. However, Muir and Moray (1996) suggest that persistence implies a dispositional characteristic and can be investigated as such. Therefore, persistence is conceptualized as a dispositional tendency to trust, specifically to trust automation. This construct will be further discussed as "automation-induced complacency" in the following sections.

Competence: Competence will be conceptualized as the extent to which the automation produces correct and appropriate recommendations or actions.

Responsibility: As conceptualized by Muir and Moray (1996), responsibility refers to the appropriateness of the system's design in comparison to alternative possible designs. However, because they found that the variance accounted for by responsibility

over and above competence was non-significant, we will adopt the conceptualization offered by Muir (1987). Responsibility will be conceptualized as the degree to which the automation provides the user with information about its motives and processes.

Predictability: Predictability will be conceptualized here as the extent to which, given the same input, the automated system will produce the same output.

Dependability: Dependability refers to an attribution that one makes that another person can be consistently relied upon (Rempel, et al., 1985). In the case of a machine, dependability will be conceptualized as the extent to which the machine is able to perform its intended function. Therefore, a machine that was not dependable would have a tendency to break down or to consistently produce error messages.

Faith: Muir (1994) suggests that faith is evident when operators use automation under conditions where there is not absolute certainty that the automation will function correctly. In situations of social trust between humans, faith is proposed to develop only after an extended period of time and is expected to emerge only after predictability and dependability have been established (Rempel, et al., 1985). However, Muir and Moray (1996) found that faith instead was the first component to emerge in the human-machine trust relationship. This finding is not counter-intuitive if one considers that when the human operator engages the automated system for the first time, he or she has no information regarding predictability or dependability on which to base trust. Instead, he or she must take a "leap of faith" that the automation will function correctly. In addition, automated systems that never produce errors or incorrect recommendations are extremely rare. Operators are, over time, bound to experience some evidence that the automated system may not be predictable and dependable, and operators may differ in the extent to

which they acknowledge this evidence. We conceptualize faith as the extent to which operators are *insensitive* to cues that the automation lacks predictability and dependability.

Complacency Potential. Some authors have suggested that people have a general propensity to either trust or not trust automated systems in general. As previously discussed, Muir and Moray (1996) conceptualize persistence as an individual difference variable related to a general propensity to trust. Parasuraman, Molloy and Singh (1993) examine a construct which similarly reflects a propensity to trust – specifically, a propensity to trust automated systems. They propose that individual differences exist in "automation-induced complacency potential," where complacency is defined as "a psychological state characterized by a low index of suspicion" (Wiener, 1981, p. 117). Parasuraman, et al. (1993) believe that some people are more likely than others to trust automated systems to the extent that they cease to monitor the system for faults. Automation-induced complacency has been implicated in numerous aircraft incidents in which pilots failed to recognize or respond to a fault in one or more automated flight systems (e.g., Danaher, 1980; Wiener, 1985). However, until the past decade there have been few efforts to empirically measure complacency or complacency potential.

Singh, Molloy and Parasuraman (1993) developed the Automation-Induced Complacency Potential Rating Scale (AICPR) in order to assess complacency potential. This scale asks respondents to rate whether they feel more comfortable using, for example, an ATM versus a human bank teller or a computer-aided library search versus a card catalog search. Because it examines generalized trust in all automation, this scale is likely to tap the construct of past experience with automated systems. While not

explicitly stated by the authors, they seem to implicitly assume that positive past experiences with automated systems in general is associated with trust in a particular automated system, which may or may not be trustworthy.

It is worth noting that the concept of complacency potential somewhat contradicts the finding of Lee and Moray (1994) that trust in one automated system does not directly affect trust in another automated system. Research is needed to examine the extent to which trust generalizes from one automated system to another.

Affect Toward the Automation. A second individual difference of interest is affect, or liking, for the specific automated system. The implementation of an automated system in the workplace can at times invoke strong affective reactions, which may affect user behaviors. For example, research on automation implementation in the workplace suggests that when workers feel that their job security may be threatened by the introduction of an automated system, they can feel negatively toward the system (Chao & Kozlowski, 1986). Research suggests that when a user has negative affect toward an automated system, he or she may avoid using it or take actions to undermine it (Sullivan, 1982). Therefore, affect toward the automation in question may influence decisions regarding automation use.

Cognitive load. People have a tendency to minimize cognitive effort (Sharit, 2003). This tendency is due to the fact that humans have limited information processing capacity. Because we cannot function well when this capacity is expended, we attempt to decrease the amount of effort we need to expend on any single task. This tendency may affect the extent to which people are motivated to rely on automated systems.

While automated systems are usually designed to decrease the amount of cognitive effort one must spend on a task, the use of automation always incurs cognitive costs (Sharit, 2003). We must learn to operate the system and continuously re-learn to keep pace with upgrades. We must then expend cognitive effort on deciding whether or not to use the automation, on going through the process of automation use, and on monitoring the system for errors. If any errors arise, we must then determine how we can correct the error. It is therefore possible for the cognitive costs of using an automated system to outweigh the cognitive benefits of using the system (Parasuraman & Riley, 1997). These cognitive load considerations may affect the likelihood that one will use an automated system. If use of the system increases cognitive load, the system is less likely to be engaged. Alternatively, if use of the system decreases cognitive load, the operator should be more likely to use it when he/she desires to decrease cognitive load.

Although cognitive load considerations may affect automation use, cognitive load has been predominantly investigated in relation to automation monitoring. Specifically, attention has focused on attempts by users to reduce their experienced cognitive load by reducing the amount of cognitive attention focused on monitoring the automated system. Research has supported the suggestion that users may fail to monitor the automation if doing so reduces cognitive load. For example, Dijkstra (1999) found that participants who consistently agreed with incorrect system advice reported using less cognitive effort than did participants who disagreed with the system at least once. Parasuraman, et al. (1993) found that when participants worked under low workload conditions, they monitored system performance accurately and identified all system errors. However, when participants were placed under conditions of high cognitive workload, they reduced

cognitive load by ceasing to monitor the automated system for errors, and in this condition they missed many system errors. Dzindolet, et al. (2001) found that participants expended less cognitive effort as automated aid reliability increased. These findings suggest that users attempting to minimize cognitive effort may not sufficiently monitor faulty automation, leading to unconditional acceptance and negative consequences.

Dispositions

I next turn to constructs related to the human user's stable dispositions.

Self-efficacy. To say that trust alone predicts behavioral choices in automation use represents a view too simplistic to accurately describe complex human behavior. As previously discussed, studies have shown that user reliance on automation does not consistently correspond to changes in the user's trust in the automation (Lee & Moray, 1992). Lee and Moray propose that task self-efficacy might be an intermediate variable in the relationship between trust and use of automation.

A widely-cited study by Lee and Moray (1994) examined the interacting effects of trust and self-confidence in operators' decisions to either place a task under automatic or manual control. Those authors defined the term "self-confidence" as "anticipated performance during manual control" (p. 154). This definition seems to overlap strongly with the psychological construct of "self-efficacy," which is defined by Bandura as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, pp. 3). Both self-confidence and self-efficacy reflect a person's beliefs about his or her ability to perform a task at a desired level. Therefore, I will hereafter use the term "self-efficacy" as synonymous with "self-

confidence." Both terms are used in reference to the user's perceptions of his or her ability to perform the task manually; they do not refer to perceptions of ability to correctly operate the automated system.

Lee and Moray (1994) used the PASTEURIZER task from Lee and Moray (1992), in which participants operated various functions in a partially-automated pasteurizing plant. Operators were responsible for maintaining acceptable performance on three functions – the feedstock pump, the steam pump, and the steam heater. The participants could choose to place each of these pumps under either automatic or manual control at any time during each trial. After participants became accustomed to operating the plant systems, two types of errors were introduced, both of which affected the feedstock pump. One type of error affected manual control of the pump, while the other affected automatic control of the pump. Therefore, during the trials in which the error affected manual control, participants were expected to allocate control to the automatic system, and vice versa.

The authors did not use existing scales of trust and self-confidence in this study, nor did they specify precise definitions of the terms. They allowed the participants to use their individual definitions of trust and self-confidence. The two constructs were measured on subjective scales and were framed to refer directly to the feedstock pump, the steam pump, and the steam heater individually.

Results on the subjective rating scales indicated that when errors occurred in the manual control of the feedstock pump, operator self-efficacy decreased and trust in the automatic system increased. Accordingly, after an error occurred while the pump was under manual control, participants tended to rely more heavily on the automation than

they had before the error. In addition, while allocation strategies varied somewhat between participants, those who chose to allocate tasks mostly to the automation generally had more trust in the automation than self-efficacy and those who mostly retained manual control generally had more self-efficacy than trust.

Lee and Moray also correlated the average difference between trust and self-efficacy for each pump with allocation to the automated system for that pump. They found significant relationships between the trust and self-efficacy difference (T – SE) for each function and automation use for that same function. The results present a strong case for the balance of trust and self-efficacy in the determination of the choice to use automation.

However, participants' choices regarding automation use did not always correspond with the difference between trust and self-efficacy. At times, participants opted to use manual control even when their trust exceeded their self-efficacy. This tendency was more pronounced in the earlier part of the study, with instances in 25% of the early trials and 13% of the later trials. The authors suggest that these deviations may be due to exploratory behavior – that is, behavior designed to "see what will happen" if the participant pursues various courses of action. While this is a possibility, it is also plausible that the deviations could be the results of unaccounted-for variables. Perhaps some of the deviations can be explained by personality variables.

Personality. Parasuraman and Byrne (2003) cite research by Parasuraman (1976) and Davies and Parasuraman (1982) that suggests that some people are more diligent monitors of automation than are others. These studies have shown that general intelligence, aptitude, reasoning skills, and memory are generally poor predictors of

monitoring performance. Personality may be a more fruitful avenue for monitoring research. Much psychological research has recently incorporated the Big Five personality traits. Two of these traits, extraversion and conscientiousness, seem particularly relevant to automation monitoring.

Conscientiousness. Conscientiousness is comprised of the tendencies to be dutiful, scrupulous, hardworking, and ambitious (McCrae & Costa, 1987). These are qualities that are likely to be important in the process of monitoring automation. Individuals high in conscientiousness are likely to be very concerned that the task is performed correctly and should be more likely to monitor the automation in order to ensure correct performance.

Extraversion. Those who are high scorers on the extraversion factor tend to be sociable, cheerful, and assertive. Introverts, on the other hand, tend to be more withdrawn and to draw energy from being alone (McCrae & Costa, 1987). Most relevant, however, is the tendency for extraverts to exhibit more sensation-seeking behavior than do introverts. Parasuraman and Byrne (2003) state that some studies have found moderate correlations between introversion and monitoring performance and suggest that this correlation may be due to introverts' lesser need for stimulation. Introversion, therefore, may be positively related to automation monitoring.

Hypotheses

See Figure 1 for an integrated model of the hypotheses. As discussed previously, persistence will be conceptualized in the human-computer trust relationship as Parasuraman, et al. (1993)'s "automation induced complacency potential." Complacency potential represents the degree to which people trust automated systems in general. Therefore, participants with a high complacency potential – and therefore high generalized trust in automation – are hypothesized to trust the specific system used in this experiment more than will those with low complacency potential. In addition, because people high in complacency potential have a dispositional tendency to trust machines across all situations, even without experience with those machines, complacency potential is hypothesized to relate positively to faith. Finally, people high in complacency potential are hypothesized to be less likely to monitor the automation's functioning than are those low in complacency potential.

H1a: Complacency potential will be positively related to subjective ratings of trust.

H1b: Complacency potential will be positively related to faith.

H1c: Complacency potential will be negatively related to automation monitoring.

As discussed by Lee and Moray (1994), the decision to use automation is thought to depend on both trust in the automation and on one's own task self-efficacy. If the user trusts the automation to perform the task correctly, s/he will be more likely to use the automation than if s/he has little trust in the automation. However, trust in the automation is balanced by the user's beliefs about how well he or she can perform the

tasks unassisted. When trust is high and self-efficacy is low, the operator will have a stronger tendency to engage the automated system, and vice versa.

H2: Self-efficacy will moderate the relationship between trust and automation use.

Furthermore, the trust – self-efficacy relationship should also affect the likelihood that a user will monitor the automation for errors. When trust is high and self-efficacy is low, operators are likely to feel that the automated system "knows best," and they will be less likely to monitor the system. However, when trust is low and self-efficacy is high, operators will be more likely to monitor the automated system because they see themselves as being equals or superiors to the automation in terms of task performance.

H3: Self-efficacy will moderate the relationship between trust and automation monitoring.

The user's desire to decrease his or her cognitive load may also affect both use and monitoring of the automated system. The degree to which use of the automation increases or decreases cognitive load may depend on the design of the system itself. When an automated system is difficult to use, the operator might experience a higher level of cognitive load when using the system than when not using the system. However, the automated system designed for the present study is relatively easy to use, and therefore, desire to decrease cognitive load should be positively related to automation engagement.

H4a: Desire to decrease cognitive load will be positively related to automation use.

Regardless of system design, monitoring the automation should always require more cognitive effort than not monitoring the automation. Therefore, it is predicted that

H4b: Desire to decrease cognitive load will be negatively related to automation monitoring.

Hypothesis 5 concerns affect toward the automated system. When affect toward the system is positive, users should be more likely to use it. Those users who experience negative emotion toward the automated system will be more likely to avoid contact with it. Therefore,

H5: Positive affect toward the automation will be positively related to automation use.

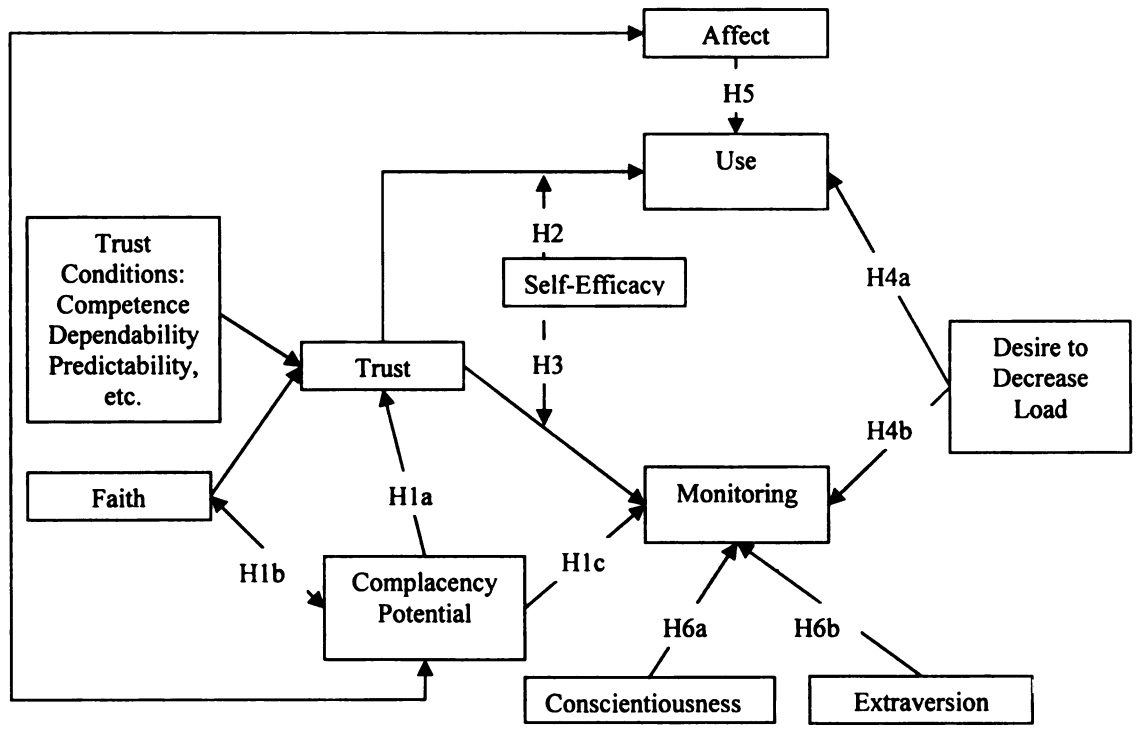
As discussed previously, it has been suggested by Parasuraman and Byrne (2003) that extraversion may be negatively related to automation monitoring. In addition, because it reflects a tendency to be diligent and careful, conscientiousness may exhibit a positive relationship with automation monitoring.

H6a: Conscientiousness will be positively related to automation monitoring.

H6b: Extraversion will be negatively related to automation monitoring

Figure 1

Experimental Model



Task Overview

In order to address the hypotheses, a task was designed that was somewhat familiar to our participants and that would be seen as important – the task of inspecting luggage. One hundred and twenty x-ray images of suitcases were developed. Some of these images contained weapons; most did not. Participants were to inspect each image as accurately and quickly as possible. After inspecting each image, participants indicated whether they would "search" the bag (they believed it might contain a weapon) or "clear" the bag (they believed it contained no weapon). In addition, they were told that an automated machine with visual capabilities was available to assist them in the inspection. This machine was termed the "Automatic Weapons Detector," or AWD. Example

images of suitcases are displayed in Figures 2, 3, and 4, and an example screenshot from the task is displayed in Figure 5. The primary focus of this study was on participants' decisions regarding this machine: their choices to use the automated system and their willingness to monitor its decisions.

Figure 2

Example X-ray Slide #1

Figure 3

Example X-ray Slide #2

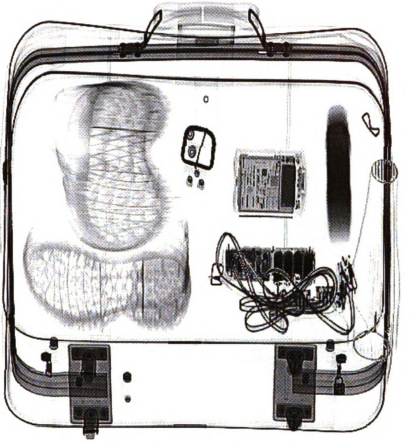


Figure 4

Example X-ray Slide #3

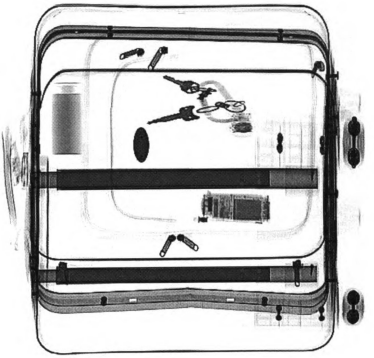
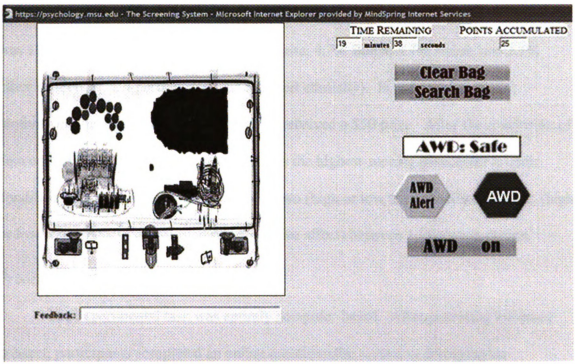


Figure 5

Screen Capture from X-ray Task



Method

Participants and Design

A sample of 253 undergraduate students from a large Midwestern university volunteered to participate in this study in return for course credit. Participants' mean age was 19.25 years. The sample was 84.9% White, 4.7% Black, 4.7% Asian, and 3.1% other ethnicities (2.4 percent declined to report ethnicity). In order to increase motivation, the top scorer in each condition received a $50 prize. After the conclusion of data collection, the $50 prize was awarded to the highest-scoring participant in each condition. Hypotheses were tested using a two (high or low trust conditions) X two (high or low self-efficacy) X two (neutral or positive affect) between participants design.

Procedure

The experimental task was entirely computer-based. After providing informed consent, participants completed an online questionnaire assessing demographic characteristics, automation-induced complacency potential, extraversion, and conscientiousness.

Next, participants completed a computer-based training session for the task of baggage screening. A transcript of this training can be found in Appendix A. The session included instructions regarding what the participants should look for (guns and knives), how to "search" or "clear" a bag, and the details of the scoring system for correct and incorrect decisions. After receiving these instructions, participants completed a one-minute practice trial during which they did not have access to the AWD. During the trial, participants received false feedback about their performance. Each participant was randomly assigned to receive high performance feedback or low performance feedback in

order to manipulate self-efficacy. This manipulation is discussed more fully in the "manipulations" section (see page 36). Following this practice trial, participants completed measures of self-efficacy and desire to decrease cognitive load.

In Training Session 2, participants received instructions on how to operate the Automatic Weapon Detector (AWD). The transcript of this session can be found in Appendix B. The training information included descriptions of the predictability, dependability, and competence of the AWD's performance. Following this training session, participants viewed a one-minute demonstration of the AWD. During this trial, participants witnessed the machine's tendency to make mistakes, to break down, or to function correctly. Following Training Session 2, participants completed measures of pre-task trust and faith.

After the two training sessions, the full task began. In this task, participants were given a time period of 20 minutes in which they were required to "screen" as many bags as possible with the fewest errors (see Figures 2, 3, and 4 for example X-ray slides). The base rate for weapon presence in this study was 30%. Airline security personnel often face time pressure in screening bags, as long lines cause customer dissatisfaction. In order to simulate these conditions, a point penalty was imposed for each minute passed during the screening task (see Table 4). The point system was pilot tested prior to data collection in order to ensure that variance was obtained in user behaviors. In sum, the point system devised for this screening task rewarded both speed and accuracy.

Following the task period, participants completed measures for post-task trust, affect, and perceived cognitive load. Finally, participants experienced a debriefing period before being dismissed from the experimental session.

Table 4

X-Ray Screening Task Scoring System.

Action	Points
Correctly Identify a Gun or Knife	+ 100 points
Miss a Gun or Knife	- 100 points
Correctly Clear a Bag	+ 25 points
Open a Bag and Find Nothing	- 25 points
Each Minute Elapsed	- 10 points

Experimental Manipulations

Trust conditions. Conditions were manipulated in order to encourage trust in the automated system to be either high or low. Based on the literature review of trust presented previously, trust conditions were manipulated by providing participants with explicit information about the predictability, competence, and dependability of the AWD. The descriptions below provide specific values for variables manipulated.

Predictability. In the high trust condition, participants were told that the AWD was 100% predictable – that is, when given the same bag to screen multiple times, the AWD would produce the same recommendation every time. In contrast, participants in the low trust condition were told that the AWD may produce different recommendations if the same bag were scanned multiple times.

Competence. Participants in the high trust condition were also told that the machine was highly competent in that it would make correct recommendations 85% of the time. In the low trust condition, participants were told that the AWD produced a

correct recommendation only 65% of the time. These percentages were pilot tested for effectiveness and were reflected in the AWD's actual functioning throughout the trials.

Dependability. In the high trust condition, participants were told that the AWD does not break down. Participants in the low trust condition were informed that the AWD is not highly dependable and may fail to operate at times. During the task trial, the AWD produced error messages on an average of 25% of bags screened.

Responsibility. Responsibility was not manipulated verbally, but was instead reflected solely in the AWD's operating procedure. In the high trust condition, the AWD provided increased levels of information about what it was doing during the 5 seconds required for it to operate. During the first 1/3 of the time, a message indicated that the AWD was "acquiring." During the second 1/3 of the time, the message indicated that the AWD was "scanning," and during the final 1/3 of the time, the message indicated that the AWD was "processing." In the low trust condition, only a single message, "scanning," was displayed for the entire 5 seconds. Therefore, in the high responsibility condition, participants were provided with more information about what the AWD was doing than they were provided in the low responsibility condition.

Self-efficacy. Self-efficacy was manipulated using the feedback provided to the participants during their practice session. Following the initial set of slides, participants in the high self-efficacy condition were told that they had been correct on 90% of their practice trials, while low self-efficacy participants were told that they had been correct on only 45% of trials.

Affect. Affect toward the AWD was manipulated using word choice in the training session. Reeves and Nass (1996) found that when a computer was labeled as a

"teammate," participants reported significantly higher liking and affect toward it than when it was simply labeled "a computer." Therefore, in the high affect condition, the training protocol referred to the AWD as "your automated assistant," whereas in the neutral affect condition, the AWD was referred to as "the Automatic Weapon Detector (AWD)."

Measures

Complacency potential. Complacency potential was assessed using the Automation-Induced Complacency Potential Rating Scale (Singh, Molloy, & Parasuraman, 1993). All items can be found in Appendix C.

Automation use. The extent to which participants made the decision to use the AWD was assessed by the proportion of total bags that were screened using the AWD. A bag was considered screened using the AWD if the participant activated the AWD prior to making a decision about whether to search or clear a bag.

Monitoring. The extent to which the participant monitored the AWD's performance was assessed using the quotient of the number of times in which the participant correctly disagreed with the AWD's recommendation by the total number of trials during which the AWD was engaged and produced an incorrect recommendation. The total number of times the participant disagreed with the AWD was also examined.

Predictability, dependability, and faith. These three trust components were measured using an adaptation of the scale developed by Rempel, Holmes, and Zanna (1985). All items can be found in Appendix D.

Competence and responsibility. Participants' perceptions of the AWD's competence and responsibility were each assessed using items created by the author for this study. See Appendix E for items.

Conscientiousness and extraversion. These dispositional characteristics were measured using the appropriate scales of the IPIP measure (Goldberg, 2001). All items can be found in Appendix F and G.

Trust. Participants' subjective trust in the AWD was measured with items created for the purposes of this study. Scale items can be found in Appendix H.

Desire to decrease cognitive load. Participants' desire to decrease cognitive load was measured immediately after the training session using items created for this study. To view the items, see Appendix I.

Manipulation Checks

Self-efficacy. Self-efficacy was assessed immediately following the training session using an adaptation of Toney and Kozlowski's (1999) self-efficacy scale. See Appendix J for items.

Affect. Participants' affect toward the AWD was assessed using self-report measures developed by the author for use in the present study. See Appendix K for items.

Perceived cognitive load. The AWD was designed with the intention of making the screening task easier. In other words, the AWD was intended to decrease participants' cognitive load on the screening task. However, it is possible for an automated system to increase the cognitive load required on a task (Sharit, 2003).

Therefore, perceived cognitive load was measured using the NASA Task Load Index (Hart & Staveland, 1988). TLX procedures can be found in Appendix L.

Results

Organization of the Results

The results begin with a discussion of the manipulation checks, followed by a presentation of scale descriptive statistics, including scale means, standard deviations, reliabilities, and intercorrelations. Next, the results of all hypotheses will be presented as well as the results of structural equation modeling tests of model fit. Before turning to the results, I will present four notes concerning the analyses.

First, it is important to note that in spite of indications to the contrary from pilot testing, the initial data collection yielded low base rates for use and disagreements with the automated system. Therefore, a mandatory use condition was added. One-half of the participants completed the revised procedure. In the revised procedure, the AWD activated automatically as each slide was presented, forcing these participants to use the AWD on every slide viewed. Independent samples t-tests revealed no significant differences between participants in the two procedure types on demographic characteristics or predictor variables. A significant difference in affect was found between groups; the mandatory AWD group had higher affect toward the AWD (see Table 5). This difference in affect (which was measured after participants had completed the full task), was likely due to the different levels of AWD use between the two groups rather than to dispositional differences. Based on the results of these t-tests, Hypotheses 2, 4a, and 5 were analyzed using participants in session Type A (AWD optional) only. Since there was little or no variance in monitoring behavior under the AWD optional condition, Hypotheses 3, 4b, 6a, and 6b were analyzed using participants in session Type

B (AWD mandatory) only. The remaining hypotheses were analyzed using all participants.

Table 5

T-tests for Significant Differences Between Session Types A and B

	t	*p*
Automation-Induced Complacency Potential	-.64	.53
Conscientiousness	.38	.70
Extraversion	-.74	.46
Desire to Decrease Cog Load	-1.57	.12
Self-Efficacy	.47	.64
Pre-task Trust	.20	.84
Faith	-.20	.84
Trust X Self-Efficacy	.03	.97
Affect	-2.51	.01
Post-task Trust	.90	.37
Automation Use	-48.62	<.01
Automation Monitoring (disagreements)	-6.49	<.01

Second, the data for AWD use exhibited a significant positive skew (see Figure 6). To correct this problem, a log transformation was performed on the data to normalize the distribution. In this transformation, eighteen participants who did not use the AWD at all on the task were removed. The transformed distribution is displayed in Figure 7. Hypotheses tested before and after the logarithmic transformation on the data did not significantly differ. Therefore, the analyses using uncorrected data are reported here.

Third, monitoring of the AWD was operationalized in multiple ways. The first operationalization was the proportion of slides on which the participant disagreed with the recommendation of the AWD.

$$M = \frac{disagreements}{slidesviewed}$$

Figure 6

The original distribution of AWD USE:

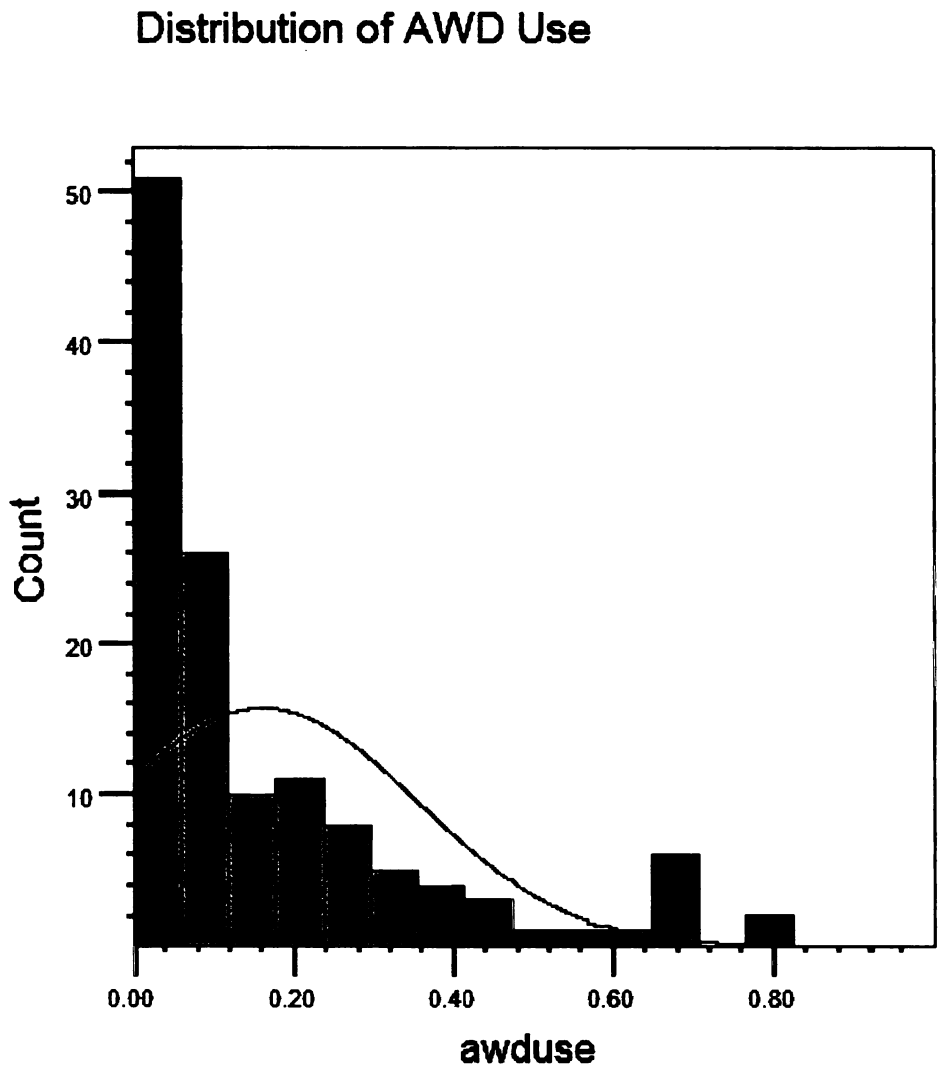
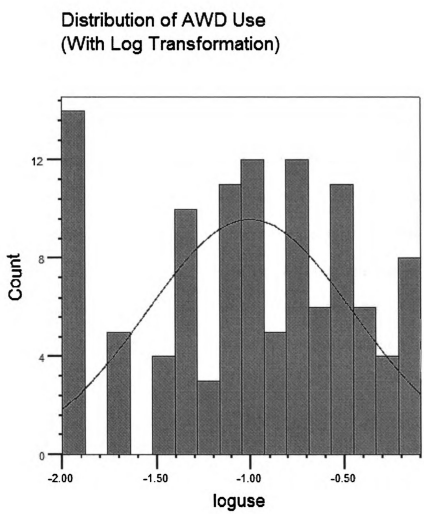


Figure 7

Transformed distribution of AWD use



The second operationalization of automation monitoring focused not on total disagreements, but on *correct* disagreements only. This second operationalization assessed not monitoring behavior in general, but only correct monitoring behavior. This

construct was created by dividing the number of disagreements made with the AWD by the number of slides on which the AWD made an incorrect recommendation.

$$M_2 = \frac{disagreements}{AWDincorrect}$$

Throughout this section, the results concerning both total disagreements and correct disagreements are presented.

Finally, subjective trust in the AWD was measured both before and after the major task was performed. Because it was expected that levels of trust would change over time as participants gained more experience with the task, hypotheses regarding subjective trust were tested using both the pre-task and post-task ratings of subjective trust (refer to appendix H for the pre-task trust items and appendices D and E for the post-task trust items). For the AWD-optional participants, the correlation between pre-task trust and post-task trust was .52, while for AWD-mandatory participants the correlation was .35. These correlations indicate that the trust level did change throughout the task.

Descriptive Statistics

Table 6 presents the means, standard deviations, minimums, and maximums for each scale used in the present study. Internal consistency reliabilities (coefficient alphas) and scale intercorrelations are found in Table 7. All scales achieved acceptable levels of reliability with the exception of the automation-induced complacency potential scale.

Table 6

Scale Ranges, Means, and Standard Deviations

	N	Min	Max	Mean	Std Dev
Complacency	253	2.5	4.83	3.50	0.37
Conscientiousness	252	2.20	5.00	3.66	0.50
Extraversion	253	2.30	5.00	3.75	0.51
Desire to Decrease Load	253	1.00	5.00	3.28	0.72
Self-Efficacy	253	2.00	5.00	3.35	0.53
Trust (pre)	253	1.33	5.00	2.89	0.63
Faith	253	1.29	4.43	2.97	0.53
Trust X SE	253	3.19	25.00	9.75	2.92
Affect	253	1.00	5.00	2.96	0.93
Trust (post)	252	1.17	4.28	2.57	0.61

Manipulation Checks

Tables 8 and 9 display the results of the manipulation checks for trust, self-efficacy, and affect. The checks were performed by correlating the experimental condition (high or low for each construct) with the individual item scores for trust (Table 8) and with the measured construct scale scores for self-efficacy and affect (Table 9). Results indicate that the manipulations for trust and self-efficacy were linked with significant differences in measured trust and measured self-efficacy. However, the manipulation for affect was non-significant.

Table 7

Scale Intercorrelations

		1	2	3	4	5	6	7	8	9
1	Complacency Potential	**.65**								
2	Conscientiousness	.12	**.83**							
3	Extraversion	.21*	.20*	**.85**						
4	Pre-task Trust	.23*	.04	.08	**.88**					
5	Self-Efficacy	.24*	.19*	.17*	.17*	**.82**				
6	Affect	.06	-.04	-.04	.31*	-.06	**.90**			
7	Post-task Trust	.07	.00	-.04	.26*	-.09	.27*	**.96**		
8	Desire to Decrease Load	-.08	-.05	-.09	-.06	-.44*	.05	.08	**.76**	
9	Faith	.36*	.05	.08	.61*	.16*	.34*	.46*	-.05	**.70**

* = *significant at* $p < .01$

$N = 253$

Note: Scale reliabilities are displayed in the diagonal of the matrix

Hypothesis Testing.

Hypotheses 1a through 1c.

Hypotheses 1a and 1b proposed that automated-induced complacency potential would be positively related to subjective ratings of trust and faith. As expected, complacency potential exhibited a significant and positive correlation with subjective pre-task trust ($r = .23, p < .01$), indicating that complacency potential, which is conceptualized as a generalized propensity to trust machines, significantly related to trust in this specific machine (the AWD) prior to the task. Therefore, Hypothesis 1a was supported. However, it is interesting to note that complacency potential failed to relate significantly to post-task trust ($r = -.02, p = .71$). Therefore, automation-induced complacency potential seemed to affect trust in an automated system only before participants had actual experiences with the automated system.

Table 8

Trust Manipulation Check:
Means and Standard Deviations of Faith and Pre-task Trust by Trust Condition and Correlations of Trust Condition with Faith and Pre-task Trust

	Faith	Pre-Task Trust
High Trust		
Mean	2.89	2.73
SD	0.55	0.62
N	133	133
Low Trust		
Mean	3.06	3.08
SD	0.48	0.59
N	120	120
Correlation†	.17*	.27**
d	.35	.59

† *Note: Point-Biserial Correlations of Trust Condition with Faith and Pre-Task Trust*
*$p = .01$
**$p < .01$

Table 9

Self-Efficacy Manipulation Check:
Means and Standard Deviations Reported Self-Efficacy by Self-Efficacy Condition and Correlation of Self-Efficacy Condition with Reported Self-Efficacy

	Self-Efficacy Condition
Low Self-Efficacy	
Mean	3.20
SD	0.49
N	131
High Self-Efficacy	
Mean	3.50
SD	0.54
N	122
Correlation†	.28*
d	.56

† *Note: Point-biserial correlation*
*$p < .01$

Table 10

Affect Manipulation Check:
Means and Standard Deviations Reported Affect by Affect Condition and Correlation of Affect Condition with Reported Affect

	Affect Condition
Neutral Affect	
Mean	2.87
SD	0.91
N	117
Positive Affect	
Mean	3.04
SD	0.94
N	136
Correlation†	.09
d	.18

† *Note: Point-biserial correlations*

Hypothesis 1b suggested that complacency potential would be significantly related to faith (measured prior to the task). This hypothesis was supported ($r = .26$, $p < .01$). The support for Hypothesis 1b suggests that those with higher automation-induced

complacency potential were more likely to be trusting of an automated system in conditions of uncertainty than were those low in automation-induced complacency potential.

Hypothesis 1c proposed that automation-induced complacency potential would be negatively related to the extent to which participants monitored the automated system. Neither measure of monitoring supported Hypothesis 1c. Total disagreements ($r = .00$, $p = .97$) and correct disagreements ($r = -.06$, $p = .52$) with the AWD were unrelated to the measure of automation-induced complacency potential. Hypothesis 1c was not supported.

One possible explanation for the lack of support found for Hypothesis 1c is that the low reliability of the complacency potential measure may have led to decreased correlations with monitoring behavior. Another possible explanation for these findings is that people, despite their initial predispositions to trust machines, were able to quickly adjust their attitudes when they gained actual experience with a specific automated system. This idea is supported by the results of Hypothesis 1a, in which complacency potential related significantly to pre-task trust, but not to post-task trust. To further examine this effect in relation to monitoring behavior, the correlations between complacency potential and monitoring were contrasted for the first half of slides and the second half of slides viewed. Results showing a stronger relationship in the first half than in the second half would support the hypothesis that experience with an automated system negates the effects of complacency potential. However, the opposite result was found. The correlation between complacency potential and total disagreements with the AWD increased in the second half when compared to the first half. Therefore,

participants either immediately adjusted their perceptions of the AWD – perhaps during the training session prior to the task – or automation-induced complacency potential might exhibit no true relationship with automation monitoring.

Hypothesis 2

Hypotheses 2 and 3 predicted that self-efficacy would moderate the relationships between trust and use and trust and monitoring, respectively. Information related to these hypotheses can be found in Table 11.

Table 11

Regression Analyses of Trust, Self-efficacy, and their Interaction on Use, Total Disagreements, and Correct Disagreements

	AWD Use			Total Disagreements			Correct Disagreements		
	n	β	*p*	n	β	*p*	n	β	*p*
Pre-task Trust	129	.32	<.01	124	-.12	.20	124	-.15	.11
Post-task Trust	129	.39	<.01	124	-.54	<.01	124	-.36	<.01
Self-efficacy*	129	-.11	.20	124	-.04	.66	124	.01	.88
Self-efficacy**	129	-.02	.80	124	-.00	.99	124	.03	.76
Pre-trust X Self-efficacy	129	-.61	.33	124	-1.44	.06	124	-.46	.55
Post-trust X Self-efficacy	129	-.86	.12	124	-1.12	.05	124	.04	.95

* = *self-efficacy in the pre-task trust model*
** = *self-efficacy in the post-task trust model*

Hypothesis 2 proposed that self-efficacy would moderate the relationship between subjective trust and AWD use. Entered in the first step of the regression equation were pre-task trust and self-efficacy. The trust by self-efficacy interaction was entered in the second step of the regression analysis. The main effect for pre-task trust was statistically significant (t=3.73, β=.32 p<.01). However, neither the main effect for self-efficacy (t = -1.28, β = -.11, p=.20) nor the interaction between trust and self-efficacy (t = -.99, β = -.61, p=.33) was found to be a significant predictor of AWD use.

Hypothesis 2 was also investigated using post-task trust. In the analysis involving post-task trust, similar results were found. Post-task trust exhibited a significant main effect on AWD use ($t = 4.65$, $\beta = .39$; $p<.01$), but self-efficacy ($t = -.26$, $\beta = -.02$, $p= .80$) and the interaction between post-task trust and self-efficacy ($t = -1.57$, $\beta= -.86$, $p=.12$) were non-significant.

In order to more fully explore Hypothesis 2, the data were split into three groups. Group 1 was composed of the 30 participants with the lowest AWD use. Group 3 was composed of the 30 participants with the highest AWD use, and group 2 was composed of the middle range participants. Groups 1 and 3 were contrasted on trust, self-efficacy, and the interaction term using independent samples t-tests to compare means (see Table 12).

Table 12

T-tests for Differences in Trust, Self-Efficacy, and Their Interaction Between Low Use and High Use Participants

	Use	N	Mean	SD	t	*p*
Pre-task	Low	30	2.57	.46		
Trust	High	30	3.18	.58	-4.51	<.01
Post-Task	Low	30	2.56	.37		
Trust	High	30	3.00	.62	-3.33*	<.01
Self-	Low	30	3.40	.53		
Efficacy	High	30	3.54	.56	.63	.53
TxSE	Low	30	8.77	2.26		
	High	30	10.57	2.77	-2.75*	.01

* Equal variances not assumed

Significant mean differences between groups 1 and 3 were found on both pre-test trust ($t = -4.51$, $p < .01$) and post-task trust ($t = -3.33$, $p < .01$), with the high use group reporting higher mean trust. This finding supports an assertion that participants with high trust are more likely to use automation than are participants with low trust. There was no

significant difference between the high use and low use groups on self-efficacy (t = .63, p = .53). The interaction between trust and self-efficacy did, however, show a significant mean difference with the high use group reporting higher means (t = -2.75, p = .01). These results suggest that the interaction between trust and self-efficacy may have predictive validity for the extreme levels of AWD use. The results of Hypothesis 2 are revisited following the discussion of Hypothesis 3.

Hypothesis 3

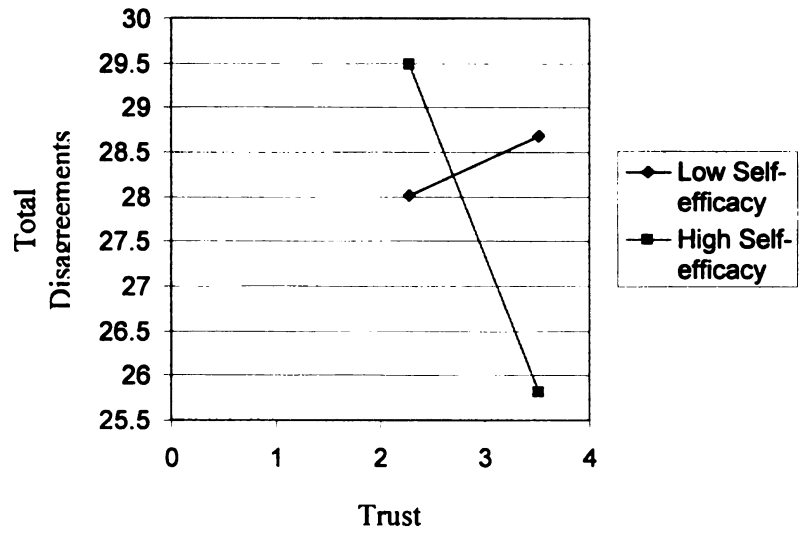
Similarly, Hypothesis 3 proposed that self-efficacy would moderate the relationship between subjective trust and monitoring. Like the relationship of self-efficacy and use, the relationship of self-efficacy and total proportion of disagreements with the machine was not statistically significant (t = -.45, β = - .04, p = .66). However, the relationship of trust with total proportion of disagreements with the automated machine also did not reach levels of statistical significance (t = -1.30, β = -.12, p = .20). The interaction between pre-task trust and self-efficacy produced a marginally significant relationship with total proportion of disagreements (t = -1.91, β = -1.44, p = .06). A graph of this interaction revealed that participants with high trust and high self-efficacy disagreed with the machine far fewer times than did participants in the other three groups (see Figure 8).

Hypothesis 3 was additionally tested using the proportion of *correct* disagreements with the machine as the monitoring form of interest. These tests yielded a somewhat different pattern of results. The direct effects of pre-task trust (t = -1.60, β = -.147, p = .11) and self-efficacy (t = .16, β = .01, p = .88) were not statistically significant. The interaction between trust and self-efficacy was also non-significant (t = -.61, β= -

.46, p = .55). However, the direct effect of post-task trust was significant (t = -4.15, β = -.36, p < .01). Self-efficacy (t = .31, β = .03, p = .76) and the interaction between post-task trust and self-efficacy (t = .07, β = .04, p = .95) were non-significant. Overall, Hypotheses 2 and 3 were not supported.

Figure 8

Graph of the Interaction Between Trust and Self-Efficacy



p = .06

There are at least two possible causes for the lack of significant effects for self-efficacy in Hypotheses 2 and 3. First, the self-efficacy manipulation may not have been sufficiently strong. The manipulation check for self-efficacy showed a significant and positive correlation between self-efficacy condition and measured self-efficacy. This result indicated that those in the high self-efficacy condition reported higher levels of self-efficacy than did those in the low self-efficacy condition. However, the magnitude of this correlation was only around r = .25. Further support for a weak manipulation is provided by the histogram of the self-efficacy distribution. A strong manipulation is

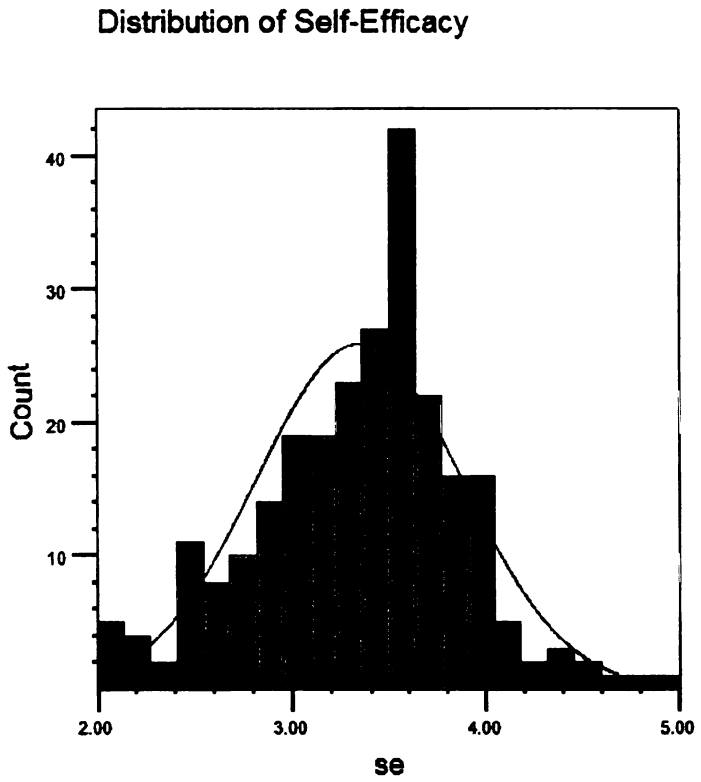
expected to yield a multimodal distribution in which two means are evident – one for the high self-efficacy condition and another for the low self-efficacy condition. However, the histogram for the data in the present study resembles a normal distribution. A normal distribution does not support the presence of a strong self-efficacy manipulation (see Figure 9). It is possible that a stronger manipulation may have yielded significant results for self-efficacy on automation use. To more closely examine the potential effects of this possibility, the participants with the highest and lowest self-efficacy were contrasted. The results showed no significant differences between the participants highest and lowest in self-efficacy for either automation use or automation monitoring. Therefore, the strength of the manipulation did not likely have a large impact on the lack of significant findings for self-efficacy.

A second possible explanation for the lack of significant results for self-efficacy on automation use concerns the duration the manipulation can be expected to last. The self-efficacy manipulation occurred early in the experimental procedure in the form of false feedback during the task training. However, participants received true feedback as they progressed through the experimental task. Therefore, the effect of the self-efficacy manipulation may have decreased steadily throughout the experimental task as participants gained a sense of their true task ability via task feedback. In order to explore this possibility, the effects of self-efficacy were tested by contrasting the halves of the experimental task. If the strength of the self-efficacy effect were decreasing throughout the experimental task, we would expect that self-efficacy would have a stronger effect on automation use on the first half of slides than on the second half of slides. However, the correlation between self-efficacy and automation use on the first half and automation use

on the second half were both non-significant and small (r = -.09, p = .30; r = -.05, p = .58, respectively). These correlations would seem to indicate that the lack of significant findings was not likely to be due to a decrease of manipulation effectiveness over time. These analyses suggest that some unknown mechanism might moderate whether self-efficacy will have a significant effect on automation use and/or monitoring.

Figure 9

Histogram of the Self-Efficacy Distribution



Hypotheses 4a and 4b

In Hypotheses 4a and 4b, desire to decrease cognitive load was proposed to be significantly and positively related to automation use and automation monitoring, respectively. Table 13 indicates that the relationships between desire to decrease

cognitive load and automation use and automation monitoring were nonsignificant. Therefore, Hypotheses 4a and 4b were not supported.

Table 13

Correlations of Desire to Decrease Cognitive Load with Use and Monitoring

	Desire to Decrease Cognitive Load		
	N	r	*p*
AWD Use	129	.11	.24
Proportion of Disagreements	124	-.06	.51
Proportion of Correct Disagreements	124	.04	.64

Hypothesis 5

Positive affect toward the automated system was proposed to predict automation use in Hypothesis 5. This hypothesis was supported ($r = .61, p < .01$). However, because affect was measured following the experimental task, the causal direction of this relationship is difficult to confirm from these findings. It is also possible that those who used the AWD more liked it more or that those who noticed fewer errors liked the machine more. To more closely examine this possibility, a hierarchical regression analysis was conducted testing whether the interaction between trust condition and AWD use significantly predicted affect. The reasoning here is that participants' affect toward the automated system will be based primarily on the AWD's functioning, and because the AWD functioned much more effectively in the high trust condition than in the low trust condition, AWD use should predict positive affect more strongly in the high trust condition than in the low trust condition. Results indicated that this was not the case. While the direct effects of trust condition ($t = 3.33, \beta = .44, p < .01$) and AWD use ($t = 7.01, \beta = 2.41, p < .01$) on affect were significant, the interaction term did not account for any significant portion of variance in affect beyond the direct effects ($t = -.35, \beta = -.121,$

p = .73). These findings indicate no significant differences in affect based on the trust condition and amount of AWD use. These results lend support to the causal direction proposed in Hypothesis 5 – that affect predicts AWD use and not vice versa.

An additional problem regarding affect concerns the possibility of multicollinearity between subjective trust and affect. Affect exhibits significant and positive correlations with faith (r = .34, p < .01), pre-task trust (r = .31, p < .01), and post-task trust (r = .74, p < .01). Although the item contents of the scales are reasonably different, the constructs nevertheless appear to be highly related. In fact, when both trust and affect (which individually show significant correlations with automation use) are entered into block 1 of the regression equation, the relationship of trust with AWD use becomes non-significant (t = 1.00, p = .32). These results indicate that multicollinearity issues may arise when research attempts to examine the simultaneous effects of both trust and affect on automation use.

Hypotheses 6a and 6b

Hypothesis 6a posits that conscientiousness will be positively related to automation monitoring. Table 14 demonstrates that the relationships between conscientiousness and total disagreements (r = .08, p = .37) and conscientiousness and correct disagreements (r = .03, p = .78) were both non-significant. In addition, contrary to Hypothesis 6b, extraversion was not significantly related to either total disagreements (r = -.03, p = .76) or correct disagreements (r = .04, p = .70).

Table 14

Correlations of Conscientiousness and Extraversion with Use and Monitoring

	Conscientiousness			Extraversion		
	N	r	*p*	N	r	*p*
AWD Use	129	.03	.72	129	.02	.83
Proportion of Disagreements	124	.08	.37	124	-.03	.76
Proportion of Correct Disagreements	124	.03	.78	124	.04	.70

Discussion

Complacency Potential Results

Hypothesis 1a proposed that automation-induced complacency potential would be positively related to trust. In partial support of Hypothesis 1a, automation-induced complacency potential was significantly related to trust before the task but not significantly related to trust after the task. This finding suggests that people did have preconceived notions about the "goodness" of automated systems, but those preconceived notions dissipated with actual experience with the automated system. This finding is a positive one in that it indicates that people can adjust their impressions of an automated system based on experience with that system.

Hypothesis 1b proposed that automation-induced complacency potential would relate positively to faith. The situation in which the participants found themselves was highly uncertain – they entered the experiment with little or no baggage screening experience and with no knowledge of the extent to which the AWD would be helpful or unhelpful on the task. In support of Hypothesis 1b, participants' level of faith (the extent to which they were willing to trust the AWD even in these conditions of uncertainty) was found to be positively related to automation-induced complacency potential. Because this finding indicates that those high in automation-induced complacency potential are more likely to trust an unfamiliar machine, it supports the conceptualization of automation-induced complacency potential as a generalized propensity to trust machines across situations.

Hypothesis 1c, which proposed significant relationships between complacency potential and monitoring behavior, was not supported. The relationship failed to reach

levels of statistical significance for both total disagreements and correct disagreements with the automated system. This finding is inconsistent with the proposition by Parasuraman and colleagues (1993) that automation-induced complacency potential is related to increased operator mistakes when working with automated systems; however, it is consistent with their qualification that significant effects will only be found on tasks on which a single operator is responsible for multiple functions. Because this was a single-task environment, the effects of automation-induced complacency potential on correct disagreements may have been suppressed.

Trust

Trust was proposed to be positively related to automated system use. This hypothesis was supported, both for pre-task trust and for post-task trust. Results indicated that the more a person trusted the automated system, the more often he or she used the automated system. This finding is consistent with the predictions of Muir (1987) and Zuboff (1988) and with the findings of Muir and Moray (1996).

In contrast, the direct effect of task self-efficacy on automated system use was non-significant. This was surprising given the evidence from prior research that those who were more confident in their ability to perform well on the screening task on their own would be less likely to use the automated system (Lee & Moray, 1994). The reason for this lack of significant effect is unclear. Supplementary analyses performed suggested that the lack of significant effect was not due to a lack of variance on self-efficacy or to a decrease of manipulation effectiveness over time.

Trust X Self-Efficacy Interaction

The interaction between trust and self-efficacy did not significantly predict either automation use or automation monitoring. In regard to automation use, supplemental analyses indicated that individuals with high self-efficacy used the AWD slightly less often than did those with low self-efficacy; however, this difference appears only at high levels of trust, suggesting that self-efficacy may only significantly interact with trust to influence automation use when the operator trusts the automated system. However, further research is needed to validate this finding.

In regard to monitoring, participants with high self-efficacy disagreed with the AWD more often than those with low self-efficacy at low levels of trust. However, that relationship flipped when trust was high, such that at high levels of trust, those with low self-efficacy are much more likely to disagree with the AWD than those with high self-efficacy. The reason for this counterintuitive finding is unclear. However, as expected, participants with high self-efficacy and low trust disagreed with the automated system more often than did participants with low self-efficacy and high trust.

When only *correct* disagreements with the machine were considered, the expected pattern emerged more clearly. Participants with low trust correctly disagreed with the AWD more often than did participants with high trust. Participants with high self-efficacy and low trust were most likely to correctly disagree with machine errors, while participants with high self-efficacy and high trust were least likely to correctly disagree with machine errors. Participants with high self-efficacy may have demonstrated both of the more extreme levels of correct disagreements because they had more confidence in their abilities to determine whether the AWD was correct or incorrect. However, because

this interaction did not achieve statistical significance, further research is needed in order to more closely examine the relationship between trust and self-efficacy.

Desire to Decrease Cognitive Load

Hypotheses 4a and 4b held that desire to decrease cognitive load would relate positively to automation use and negatively to automation monitoring, respectively. Neither hypothesis was supported in the present study. However, desire to decrease cognitive load was significantly negatively related to self-efficacy ($r = -.44$, $p < .01$). This significant correlation indicates that participants who believed that they were weaker performers on the task held a stronger desire to decrease their cognitive load on the task. This finding suggests that the desire to decrease cognitive load scale may serve as a proxy for perceived task difficulty.

Based on the research of Dijkstra (1999), it was expected that individuals who used the automated system more and disagreed with it less would experience lower levels of cognitive load. However, this was not the case in the present study. Perceived cognitive load was assessed using the NASA Task Load Index, NASA TLX (Hart & Staveland, 1988). See Table 15 for TLX score information.

Results indicated no significant relationship between AWD use and cognitive load or AWD monitoring and cognitive load. An independent samples t-test also indicated that there were no significant differences in perceived cognitive load between the two session types (optional AWD use and mandatory AWD use). These results reveal that use and monitoring of the automated system had no significant effect on cognitive load. This lack of effect could explain the lack of significant findings between desire to decrease cognitive load and automation use and monitoring. If use of the AWD did not

result in a change in perceived cognitive load, then desire to decrease cognitive load would not be expected to affect participants' AWD use decisions. Likewise, if AWD monitoring did not increase cognitive load, then desire to decrease cognitive load would not be expected to significantly predict AWD monitoring.

Table 15

Perceived Cognitive Load Minimums, Maximums, Means, and Standard Deviations

	N	Minimum	Maximum	Mean	Std. Deviation
Type A Participants					
Mental	128	2	45	22.89	9.02
Physical	128	0	12	.38	1.56
Temporal	128	1	36	12.20	8.24
Effort	128	2	45	12.56	8.47
Performance	128	0	36	14.79	8.57
Frustration	128	0	45	10.23	6.64
Total Score	128	25	105	73.02	15.77
All Participants					
Mental	252	2	45	23.75	9.39
Physical	251	0	25	.43	2.01
Temporal	252	0	36	11.40	7.87
Effort	252	0	45	13.08	8.55
Performance	252	0	36	14.69	8.33
Frustration	252	0	45	9.94	6.63
Total Score	251	25	115	73.25	16.84

It is also possible that the task design did not allow cognitive load effects to surface. Because the study took place in a controlled lab setting, participants were encouraged to give their full attention to the task. In fact, there was little else that participants *could* give attention to in the laboratory environment. In addition, even when a participant used the AWD, he or she still had to make the ultimate decision about whether to clear or search each bag. These conditions might reasonably reflect the nature of an airport security screening job, but they do not provide many opportunities for the

participant to reduce his or her cognitive load. While the AWD scanned a bag, the image of the bag was still present on the screen, and participants had little else to do but stare at the bag and formulate their own hypotheses about whether or not a weapon was present in the image. In that sense, participants did about the same amount of cognitive work (visually searching the bag and formulating a plan to either search or clear the bag) regardless of whether the AWD was activated. Therefore, it seems that task and automation design influence the extent to which cognitive load concerns will predict behavior. In a task with few other distractions, cognitive load may not play a large role in determining automated system use and/or monitoring behaviors.

Affect

Affect toward the AWD exhibited a strong positive relationship with AWD use. This finding provides empirical support for the propositions by authors such as Chao and Kozlowski (1986) that operators with negative feelings for an automated system will avoid using it. In fact, affect exhibited the strongest effect on automation use in the present study. In addition, in exploratory analysis, a negative relationship was found between affect and automation monitoring. This result suggests that when people are forced to use an automated system but feel negatively toward it, they may avoid monitoring it rather than expending extra effort on catching machine errors.

Conscientiousness and Extraversion

Past meta-analytic research has linked conscientiousness with job performance (e.g., Barrick & Mount, 1991; Hogan & Holland, 2003). Because correct disagreements with the AWD should be related to task performance in the current study, a positive relationship was expected between conscientiousness and correct disagreements with the

AWD. Contrary to expectations, conscientiousness failed to exhibit significant relationships with automation monitoring, as did extraversion. It is possible that the duration of the study was not sufficiently long to bring out differences in monitoring behavior across levels of conscientiousness.

In addition, participants may have been subject to strong motivational effects. One source of motivation for the participants may have resulted from a cross-over of real-world concerns into the experimental task. Airline safety is currently a major concern in the world outside of the study, and thus participants may have felt a strong desire not to allow any weapons to go through undetected. The students therefore may have experienced a tendency to over-search the bags. In addition, participants were told prior to the task that the top performer would receive a $50 prize, an amount that may have provided significant motivation for undergraduate students. These motivational factors may have provided a motivational situation of sufficient strength to overwhelm the effects of conscientiousness on monitoring behavior. Thus, the combination of a relatively short task and strong motivational factors may have masked any effect of conscientiousness on monitoring performance.

With respect to extraversion, the duration of the experimental task may have again played an important role in the strength of the relationships found. The studies that provided the basis for the hypothesized relationship between extraversion and automation monitoring were pulled primarily from the vigilance stream of research. Research on automation use and vigilance often consists of experimental tasks lasting for several hours (e.g., Lee & Moray, 1994; Muir & Moray, 1996). A desire for sensory stimulation may be manifested only after a longer amount of time than the 20-minute task time used

here. The experimental task in the present study lasted for a relatively brief amount of time, whereas the "vigilance decrement" – a marked decrease in performance in error detection – tends to occur over a length of time spent on task (e.g., Grier, et. al, 2003). Perhaps a longer task duration is necessary to detect any relationship between extraversion and monitoring performance. A second alternative is that the experimental task used in the present study was inherently more stimulating than were the tasks used in the vigilance studies in which significant relationships were found. A more stimulating task would decrease the sensory depravation that was hypothesized to lead to decreased monitoring performance for extraverts.

Limitations

The extent to which a sample of undergraduate college students can generalize to a broader population may be sometimes questionable. In this case, the student sample used may have had effects in two ways. First, it was assumed that all of the sample participants entered the experiment with equivalent, and negligible, X-ray screening experience. By extension, we assumed that all participants would have approximately equal, and low, self-efficacy on the task. However, this distinction between a college and organizational population was likely to have reduced impact due to the fact that we were able to significantly manipulate self-efficacy during the task training. A graph of post-manipulation self-efficacy scores revealed an approximately normal distribution of self-efficacy. This distribution suggests that while the manipulation was not extremely strong, variance on self-efficacy was achieved.

A second and more significant issue, however, is the difference in motivation between college students and professional samples. A sample of actual X-ray screeners

might be more motivated to achieve high scores on the X-ray task than might a sample of undergraduate students because actual X-ray screeners might identify with the task more strongly. In the present study, we attempted to increase motivation to perform by offering a $50 prize to the top scorer in each experimental condition. According to comments made by the students, this $50 prize seemed to provide a significant motivational incentive. While a one-time prize cannot match the level of motivation experienced on an actual job, we believe that it provided some encouragement for the students to achieve high performance.

An additional concern is that the task design may not have allowed certain hypothesized relationships to materialize. For example, the effects of conscientiousness and extraversion may require longer periods of time to develop, and cognitive load effects may only surface when people are required to balance multiple tasks. Further research is required to more fully determine the populations and tasks for which these effects are significant.

Future Research Suggestions

The current study focused primarily on person factors impacting automation use and monitoring, such as trust, self-efficacy, and personality factors. Structural equations analyses suggested that trust may mediate the relationships of individual differences with automation use and automation monitoring. In addition, the results suggested that the psychological processes leading to automation use and monitoring decisions may be highly related. However, these propositions should be subjected to future testing in light of the fact that the model modifications made here capitalized on chance.

In addition, the current study found a surprising lack of relationship between self-efficacy and automation use and monitoring decisions. The reasons for this lack of relationship are unclear but might be related to some element in task design or sample characteristics. Future research may shed light on the situations in which self-efficacy is and is not significantly related to automation use and monitoring.

While self-efficacy did not show any significant relationships with use and monitoring, affect toward the automated system exhibited strong effects with both. The results of this study suggest that affect may play a very large role in the decisions that operators make about automation use and monitoring. Antecedants of affect toward automation might be explored, as might potential mediators of the relationships between affect and the decisions made about use and monitoring.

Finally, while the present study was designed to investigate person factors affecting automation use and monitoring, the situation in which the operator finds him or herself may also impact these two decisions. Future research should identify situational characteristics that may interact with person factors to affect automation use and monitoring decisions.

Conclusion

The present study contributed to the literature on human-automation interaction through an examination of the ways in which individual differences affect automation use and monitoring. Individual differences in trust and affect exhibited effects on both automation use and automation monitoring, indicating that those high in trust or affect were more likely to use and less likely to monitor the automated system. These results for trust and affect toward the system suggest that individual differences in attitudes

toward automated machines exist and do have effects on automated system use. It is important to note that levels of trust and affect are most likely also influenced by machine characteristics such as competence and dependability. Future research might further investigate the relative influence of the four machine characteristics discussed in the present study (competence, dependability, responsibility, and predictability) and individual differences on levels of trust and affect.

In the present study, no significant effects were shown for self-efficacy, desire to decrease cognitive load, conscientiousness, or extraversion. The non-significant finding for self-efficacy was particularly unexpected because the influence of self-efficacy on automation use had received previous support in the literature (i.e., Lee & Moray, 1994). The reason for this unexpected result is unclear. The manipulation check for self-efficacy indicated a successful manipulation with a moderate effect size ($d = .56$). In addition, contrasts of the correlations between self-efficacy and automation use for the first half of the task and the second half of the task indicated that the relationship did not decay over time; instead, it seems that the relationship was non-significant even on the initial slides. Because the lack of significant effect did not seem to be caused by methodological factors, some aspect of the task, situation, or sample may be relevant to the emergence of a significant self-efficacy effect. Therefore, future research should investigate the potential moderating effects of task design (e.g., length, complexity, etc.) and sample characteristics (e.g., motivation, task experience, etc.) on the relationship between self-efficacy and automation use. Such analyses may also shed light on the potential impact of other individual differences (such as conscientiousness or desire to decrease cognitive load) that did not exhibit significant effects in the present study.

In future studies, researchers may also wish to investigate more specifically the subtypes of use and monitoring errors that operators make. It is possible that false positive errors and false negative errors may have different causes. Individual characteristics such as risk aversion may differentially affect operators' willingness to make false positive and false negative errors. In addition, if the consequences of false positives and false negatives are asymmetrical, operator behavior is likely to be affected.

Overall, the present research supports the assertion that individual differences have a role to play in the decisions users make about using and monitoring automated systems. In addition, the results suggest that it may be difficult to identify the effects of these individual difference characteristics independent of a consideration of situational and task characteristics. While this study focused primarily on the interplay of the user's individual differences and the machine's characteristics (e.g., competence, dependability, etc.), future research should also more strongly consider task characteristics (e.g., complexity, consequences for false positives versus false negatives) and situational characteristics (e.g., environmental distractions). Accurate prediction of automation use and automation monitoring decisions may depend on the interactions of all four types of influences: machine, person, task, and situation.

Further research is warranted on the effects of individual differences on automation use and monitoring decisions. Although non-significant results were found for self-efficacy, extraversion, and conscientiousness in the present study, the reasons for these findings are unclear. As previous research has suggested that relationships among these individual differences and automation use and monitoring may exist, researchers

should investigate the potential moderating effects of task characteristics on these relationships.

Relevant task characteristics may include the length of the task, the complexity of the task, and consequences for success and failure inherent in the task. Task length may be important in that as the length of the task increases, the effects of traits related to motivation may increase. For example, near the beginning of the task, all participants may be equally motivated to perform well. However, as the task wears on, the effects of conscientiousness and extraversion on use and monitoring may become increasingly evident. To investigate these effects, researchers should vary task length or contrast the relationships between conscientiousness or extraversion and automation use or monitoring in the early stages of the task versus the late stages of the task. Note that to see significant effects, the task may need to be at least 30-60 minutes in duration.

Situational characteristics may be most likely to affect use and monitoring decisions via their effects on cognitive capacity. In situations where operators are required to divide attention among multiple tasks or in environments that are distracting, cognitive load effects are more likely to be found than in cognitively "easy" situations.

Finally, research in the area of social cognition may inform future research on operator use and monitoring. As technology becomes more intelligent and more human-like, operators are more likely to respond to these technologies as they would respond to another human (Reeves & Nass, 1996). Therefore, future researchers may wish to consult the social cognition literature on topics such as impression formation and person perception in order to determine the extent to which these processes generalize to human-machine relationships.

References

Arnold, V. & Sutton, S.G. (1998). The theory of technology dominance: Understanding the impact of intelligent decisions aids on decision makers' judgments. *In J.E. Hutton (Ed.) Advances in accounting behavioral research, 1*, 175-194.

Bandura, A. (1997). *Self-Efficacy: The exercise of control.* New York: W. H. Freeman.

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

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1-26.

Bobrow, D. G., Mittal, S., & Stefik, M. J. (1986). Expert systems: Perils and Promise. *Communications of the ACM, 29*, 880-894.

Chao, G. T., & Kozlowski, S. W. J. (1986). Employee perceptions on the implementation of robotic manufacturing technology. *Journal of Applied Psychology, 71*(1), 70-76.

Connors, M. M., Harrison, A. A., & Summit, J. S. (1994). Crew systems: Integrating human and technical subsystems for the exploration of space. *Behavioral Science, 39*, 183-212.

Danaher, J. W. (1980). Human error in ATC system operations. *Human Factors, 22*(5), 535-545.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies, 58*, 719-735.

Dijkstra, J. J. (1998). User agreement with incorrect expert system advice. *Behavior and Information Technology, 18*(6), 399-411.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58,* 697-718.

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.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight deck automation issues. *The International Journal of Aviation Psychology, 9*(2), 109-123.

Goldberg, L. R. International Personality Item Pool (2001). A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences (http://ipip.ori.org/). Internet Web Site.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P. A., Ed.; Meshkati, N., Ed.; *Human mental workload*, Oxford, England: North-Holland.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*(1), 100-112.

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and ethics. *Academy of Management Review, 20,* 379-400.

IBM. (n.d.). *Deep Blue Wins 3.5 to 2.5.* Retrieved November 14, 2003, from *http://www.research.ibm.com/deepblue/home/html/b.html.*

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*(1), 53-71.

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

Lee, G., & Perry, J. L. (1999). Are computers boosting productivity?: A test of the paradox in state governments. *Center for Research on Information Technology and Organizations. I.T. in Business*. Paper 121. http://repositories.cdlib.org/crito/business/121

Leli, D. A. & Filskov, S. B. (1984). Clinical detection of intellectual deterioration associated with brain damage. *Journal of Clinical Psychology, 40*(6), 1435-1441.

Lewandowski, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied, 6*(2), 104-123.

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63,* 967-985.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52(*1), 81-90.

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

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

Neuman, P. G. (1986). AI in medical diagnosis and aviation. *ACM SIGSOFT, Software Engineering Notes, 11*(2), 5.

Parasuraman, R. & Byrne, E. A. (2003). Automation and human performance in aviation. In: Tsang, P.S. & Vidulich, M. A. (Eds.) *Principles and practice of aviation psychology*. Lawrence Erlbaum Associates, Mahwah, NJ, 311-356.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency." *The International Journal of Aviation Psychology, 3*(1), 1-23.

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

Reeves, B. & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Stanford, CA: CSLI Publications.

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

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly, 41,* 574-599.

Sharit, J. (2003). Perspectives on computer aiding in cognitive work domains: Toward predictions of effectiveness and use. *Ergonomics, 46*(1-3), 126-140.

Sheridan, T. B., & Hennessy, R. T. (eds), (1984). *Research and Modeling of Supervisory Control Behavior* (National Academy Press, Washington).

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology, 3*(2), 111-122.

Sollins, K. R. (1986). A medical risk of computers. *AGM SIGSOFT, Software Engineering Notes, 11*(3), 11.

Sparaco, P. (1994). A330 crash to spur changes at airbus. *Aviation Week and Space Technology, 141*(6), 20-22.

Stratopoulos, T., & Dehning, B. (2000). Does successful investment in information technology solve the productivity paradox? *Information and Management, 38*(2), 103-117.

Sullivan, M. J. (1982). *Managing to mismanage robot productivity programs.* Dearborn, MI: SME Technical Paper.

Toney, R. J., & Kozlowski, S. W. J. (April, 1999). Effects of evaluative feedback and task difficulty on learning and training performance. In S. Gully (Chair), Learning to fail or failing to learn? The role of errors, failure, and feedback in learning environments. Symposium presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Waterman, D. A. (1986). *A guide to expert systems.* Reading, MA: Addison-Wesley Publishing Company.

Wiener, E. L. (1981). Complacency: Is the term useful for air safety? In *Proceedings of the 26th Corporate Aviation Safety Seminar*. Flight Safety Foundation, Denver, 116-125. p. 117

Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors, 27,* 75-90.

Will, R. (1991). True and false dependence on technology: Evaluation with an expert system. *Computers in Human Behavior, 7,* 171-183.

Yu, V. L., Fagan, L. M., Wraith, S. M., Clancey, W. J., Scott, A. C., Hannigan, J., Blum, R. L., Buchanan, B. G., & Cohen, S. N. (1979). Antimicrobial selection by a computer: A blinded evaluation by infectious diseases experts. *Journal of the American Medical Association, 242(*12).

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power.* Basic Books, New York.

Appendix A

User Training Script

This is a simulation of the job of airport security screener. You'll be looking at X-ray images of luggage and deciding whether or not to search each bag. If you think you see a GUN or a KNIFE in the X-ray image, you should SEARCH the bag. If not, you should CLEAR the bag. To SEARCH a bag, click on the button labeled, "search bag." To CLEAR a bag, click on the button labeled, "clear bag."

After you search or clear the bag, feedback will be presented on the screen that tells you whether your decision was correct. This feedback will remain on the screen for a few seconds, and then you'll be presented with the next X-ray to examine.

You will gain points for correct decisions and lose points for incorrect decisions. The point breakdown is as follows:

* Find a weapon: +100 points
* Miss a weapon: -100 points
* Search a bag with no weapon inside: -25 points
* Clear a bag with no weapon inside: +25 points
* Each minute passed: -10 points

Take a minute to make sure you know how points are gained and lost. Notice that you'll lose 10 points for every minute passed, so try to work as quickly as you can without making mistakes.

Now you'll have a chance to practice screening X-rays before the main simulation begins. A small clock will tell you how much time you have left in this practice session. When time is up, we'll let you know how you did on the practice session and ask you to fill out another short survey. When you are ready to begin, click the button below.

Appendix B

Machine Training Script
(AWD Optional)
Low trust and neutral affect conditions

When you are completing the main simulation, you can use the AWD to help you out. The AWD scans each X-ray image and alerts you if it detects anything suspicious.

For each X-ray, you will have the choice whether or not you will use the AWD. To turn it on, click on the button that says "AWD on." When it is activated, a small light comes on, indicating that the AWD is operating.

If the AWD detects something it thinks is suspicious, a red light will begin to flash, and the message "ALERT!" will appear on the screen. If the AWD thinks the bag is safe, the message "SAFE" will appear in blue on the screen.

Even when the AWD is activated, you still have to decide whether to SEARCH or CLEAR the bag. That means that even if the AWD produces a recommendation, it is up to you--the operator--to make the final decision about whether to search or clear the bag.

There are two things about the AWD to keep in mind:
1.) The AWD is a complex technology and takes a few seconds to analyze each X-ray image.
2.) The AWD isn't 100% accurate. In fact, it is accurate 65% of the time. In addition, it sometimes will not be able to scan an image and will produce an error message. If that happens, you won't be able to use the AWD on that X-ray. Also, if the AWD scans the same bag multiple times, it will not produce the same recommendation each time.

PLEASE REVIEW THE INFORMATION ABOUT HOW TO USE THE AWD AND BE SURE YOU UNDERSTAND IT BEFORE PROCEEDING.

Next, you will view a DEMONSTRATION of the AWD scanning bags by itself. This demonstration will allow you to see how the AWD works before you use it yourself. In this demonstration, you don't have to click "search bag" or "clear bag." However, feedback will appear on the screen telling you whether the AWD was right or wrong on each X-ray. Now, watch the AWD.

Appendix C

Automation-Induced Complacency Potential Measure
(Singh, Molloy, & Parasuraman, 1993)

1. I think that automated devices used in medicine, such as CT scans and ultrasound, provide very reliable medical diagnosis.
2. Automated devices in medicine save time and money in the diagnosis and treatment of disease.
3. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because it is more reliable and safer than manual surgery.
4. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.
5. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.
6. Automated devices used in aviation and banking have made work easier for both employees and customers.
7. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed trap in case the automatic control is not working properly.
8. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.
9. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.

10. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.
11. I feel safer depositing my money at an ATM than with a human teller.
12. I have to tape an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my VCR rather than manual taping.

Appendix D

Post-task Predictability, Dependability, and Faith Items

(Adapted from Rempel, Holmes, & Zanna, 1985)

1. When I encounter an unfamiliar object in a bag, I would not feel worried letting the AWD screen the bag (faith)
2. I can count on the AWD to be there to help me out when I need it (dependability)
3. In general, the AWD does the same thing each time I use it (predictability)
4. The AWD has proven to be trustworthy, and I am willing to depend on it (dependability)
5. I am familiar with the functioning of the AWD, and I can rely on it to behavior in certain ways (dependability)
6. Even when I don't know how the AWD will answer, I feel comfortable letting it screen the bags (faith)
7. Though times may change and terrorists may try new plans, I know that the AWD will always be ready to help me stop them (faith)
8. I am never certain that the AWD will give me the right answer (predictability)
9. The AWD is very unpredictable. I never know if it is going to give me the correct advice (predictability)
10. I feel very uncomfortable when the AWD makes decisions that might affect how well I do my job (predictability)
11. I have found that the AWD is usually dependable, especially on important decisions (dependability)
12. The AWD functions in a very consistent manner (predictability)

13. When I use the AWD, my future screening performance is an unknown which I worry about (faith)
14. Whenever I have to make a decision about a strange looking bag, the AWD will do its best to help me make the right choice (faith)
15. I can rely on the AWD to help me make a correct decision when I can't make the decision myself (faith)
16. The AWD can be counted on (dependability)
17. I have to keep alert, or the AWD might convince me of the wrong decision (dependability)
18. I am certain that the AWD will not break down (dependability)
19. I sometimes avoid using the AWD because it is unpredictable and I fear that it will give me the wrong advice (predictability)
20. I can rely on the AWD to function at its best every time I use it (dependability)
21. Even if I were to continue working as a screener, I would never guarantee that I would still use the AWD 1 year from now (faith)
22. When I am using the AWD, I feel secure in facing unknown new situations (faith)
23. Even when the AWD gives advice that seems like it might be wrong, I am confident that it is still operating correctly (dependability)
24. I am willing to let the AWD make decisions for me (dependability)

Appendix E

Competence Scale Items

1. The AWD usually gave me the correct recommendation.
2. The AWD was generally right.
3. The AWD was very competent.

Responsibility Scale Items

1. The AWD kept me informed of what it was doing.
2. The way the AWD works is a mystery to me.
3. I understand the AWD's process.

Appendix F

Conscientiousness Scale Items

(Goldberg, 2001)

1. Am always prepared.
2. Pay attention to details.
3. Get chores done right away.
4. Carry out my plans.
5. Make plans and stick to them.
6. Waste my time.
7. Find it difficult to get down to work.
8. Do just enough work to get by.
9. Don't see things through.
10. Shirk my duties.

Appendix G

Extraversion Scale Items

(Goldberg, 2001)

1. Feel comfortable around people.
2. Make friends easily.
3. Am skilled in handling social situations.
4. Am the life of the party.
5. Know how to captivate people.
6. Have little to say.
7. Keep in the background.
8. Would describe my experiences as somewhat dull.
9. Don't like to draw attention to myself.
10. Don't talk a lot.

Appendix H

Pre-task Trust Items

1. Overall, I think the AWD is trustworthy.
2. The AWD is a competent performer.
3. I can depend on the AWD.
4. I find the AWD very predictable.
5. I have faith that the AWD to perform well.
6. The AWD is very responsible.

Appendix I

Desire to Decrease Cognitive Load Scale Items

1. This task seemed a lot harder than it should have been.
2. I wish this task had been easier.
3. I had to pay more attention to this task than I really wanted to.
4. I felt too much time pressure while screening the bags.
5. I felt myself getting more frustrated than I would have liked.

Appendix J

Self-Efficacy Scale Items

(Adapted from Toney & Kozlowski, 1999)

1. I can meet the challenges of this baggage screening task.
2. I am confident in my understanding of how information cues are related to decisions.
3. I can deal with decisions about bags under ambiguous conditions.
4. I am certain that I can manage the requirements of this screening task.
5. I believe I will fare well in this task if the required speed is increased.
6. I am confident that I can cope with this screening simulation if it becomes more complex.
7. I believe I can develop methods to handle challenging aspects of this task.
8. I am certain that I can cope with any distractions that may arise.

Appendix K

Affect Toward the AWD Scale Items

1. I liked using the AWD
2. I was glad I had the option of using the AWD to assist me
3. Overall, I feel positively toward the AWD
4. The AWD made me feel uncomfortable
5. I wish the AWD had never been created

Appendix L

Perceived Cognitive Load Measure

(NASA Task Load Index)

The TLX assesses six dimensions of cognitive load, as presented in Table 16 below.

Table 16

Dimensions of the NASA TLX Scale

Dimension	Definition
Mental Demand	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Frustration Level	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

These definitions were presented to the participants, and then participants rated the extent to which the task was high or low on likert-type scales ranging from 1 to 9.

Figure 10

Example TLX Scale

TEMPORAL DEMAND: how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?

- It had almost no effect
- It had a small effect
- It had a moderate effect
- It had a large effect
- It had an extreme effect

After participants provided ratings for each of the six dimensions, the dimensions were weighted according to their subjective importance. Importance was assessed using a series of 15 pairwise comparisons. Participants were presented with two of the six dimensions and rated which dimension was the more important one in establishing their overall workload.

Example:

Mental Demand	vs.	Physical Demand
Effort	vs.	Temporal Demand
Mental Demand	vs.	Frustration

The number of times each dimension was selected was calculated and used as a weight. The participant's Likert scale rating for each dimension was multiplied by its

weight. The total cognitive load score is the sum of the weighted scale ratings divided by 15 (the number of comparisons).

Appendix M

Correlation Matrices for Automation Use and Monitoring

Table 17

Sample Correlation Matrix for Use

	Use	Compl	Consc	Extrav	S.E.	T.xS.E.	Faith	Trust	Affect	Desire
Use	**(.19)**									
Complacency	.12	**(.36)**								
Consc	.03	.15	**(.51)**							
Extrav	.02	.34*	.24*	**(.52)**						
S.E.	-.07	.25*	.18*	.29*	**(.53)**					
T.xS.E.	.18*	.33*	.17	.29*	.66*	**(3.00)**				
Faith	.28*	.30*	.08	.20*	.13	.58*	**(.53)**			
Trust	.30*	.24*	.05	.15	.13	.82*	.69*	**(.64)**		
Affect	.61*	.13	-.06	.05	-.15	.19*	.44*	.39*	**(.90)**	
Desire	.11	-.08	-.08	-.22*	-.45*	-.28*	-.00	-.03	.10	**(.76)**

N = 129

**p < .05*

Note: Scale standard deviations are presented in the diagonal

Table 18

Sample Correlation Matrix for Monitoring

	Monitor	Compl	Consc	Extrav	S.E.	T.xS.E.	Faith	Trust	Affect	Desire
Monitoring†	**(17.25)**									
Complacency	-.06	**(.38)**								
Consc	.03	.09	**(.49)**							
Extrav	.04	.08	.16	**(.50)**						
S.E.	-.02	.23*	.20*	.04	**(.53)**					
T.xS.E.	-.12	.30*	.13	.05	.69*	**(2.84)**				
Faith	-.08	.21*	.02	-.06	.20*	.50*	**(.53)**			
Trust	-.14	.22*	.03	.01	.21*	.85*	.53*	**(.62)**		
Affect	-.38*	-.01	-.01	-.16	.05	.19*	.26*	.23*	**(.95)**	
Desire	.04	-.08	-.00	.06	-.44*	-.28*	-.11	-.10	-.04	**(.68)**

N = 123

† Correct Disagreements

**p < .05*

Note: Standard deviations presented in parentheses

Appendix N

Validation of the Automation-Induced Complacency Potential Rating Scale

This study included an effort to validate Singh, Molloy and Parasuraman's (1993) automation-induced complacency potential rating scale. This rating scale was designed to assess an individual's stable propensity to become complacent in regard to automated systems. The subscales of the complacency potential rating scale are trust, compliance, reliance, and safety. To my knowledge, the psychometric properties of these subscales had yet to be validated by experimenters other than the authors. None of the four subscales proposed by Singh and colleagues achieved acceptable levels of reliability in the present sample (see Table 19). An exploratory factor analysis was performed which did not support the four factor structure. The scree plot from this analysis is displayed in Figure 11.

Table 19

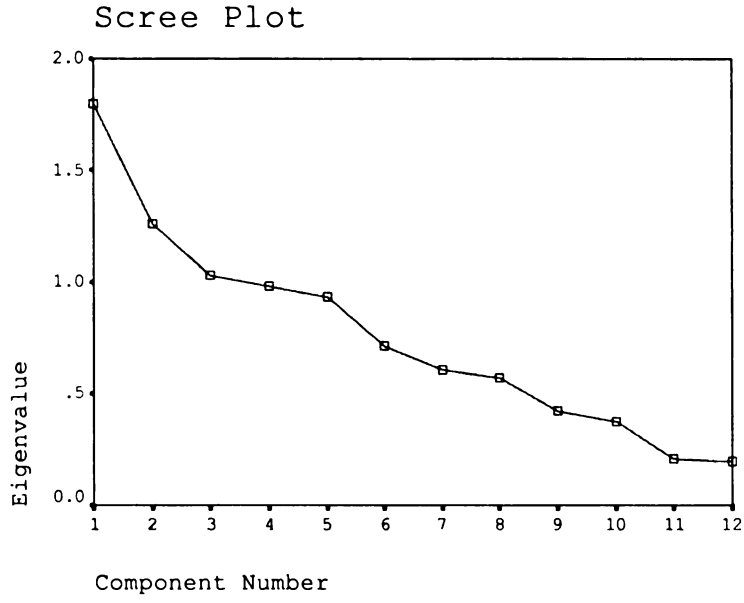
Reliabilities of the Automation-Induced Complacency Potential Subscales

	N	# of Items	Alpha
Confidence	251	4	0.55
Reliance	249	3	0.29
Trust	251	3	0.30
Safety	253	2	0.08

An examination of the item content provides a possible explanation. These subscales were developed empirically (Singh, et al., 1993), and the correspondence between the some items' content and their scale labels is unclear. For example, the item "ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people" was assigned to the "reliance" subscale as opposed to the "safety" subscale, and the item "I have to tape an important TV program for a class assignment; to

ensure that the correct program is recorded, I would use the automatic programming facility on my VCR rather than manual taping" was assigned to the "safety" subscale as opposed to the "reliance" subscale.

Figure 11
Scree Plot – Exploratory Factor Analysis for Automation-Induced Complacency Potential



Scale refinement was attempted in order to improve internal consistency. Revision based on item-total correlations resulted in a one-factor scale with seven items and a maximum coefficient alpha of α=.64, still below the acceptable level of .70 (see Table 20).

In conclusion, the present study was unable to find support for the psychometric properties of the automation-induced complacency potential rating scale, either as a whole or in terms of its individual subscales. First, no subscale achieved an acceptable level of internal consistency, as subscale alpha coefficients were .55, .29, .30, and .08, respectively. The alpha coefficient for the entire scale was .60, and the revised scale

alpha coefficient was .64 – slightly better but still below the conventionally acceptable level of .70.

Table 20

Revised Automation-Induced Complacency Potential Scale Items and Statistics

Item	Mean	Std Dev	Item-Total Correlation
I think that automated devices used in medicine, such as CT scans and ultrasound, provide very reliable medical diagnosis	4.17	.58	.30
Automated devices in medicine save time and money in the diagnosis and treatment of disease	4.01	.67	.43
If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because it is more reliable and safer than manual surgery	3.45	.87	.36
Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer	3.81	.72	.38
ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people	3.15	.95	.36
Automated devices used in aviation and banking have made work easier for both employees and customers	4.12	.56	.45
Bank transactions have become safer with the introduction of computer technology for the transfer of funds	3.25	.92	.28
	Overall Alpha		.64

Appendix O

Post-hoc Analyses of the Effects of Slide Difficulty

The X-ray slides were designed to fit into a classification of easy, medium, and challenging levels of difficulty. Easy slides contained between two and four items, and any weapons contained in those slides were easily visible. Medium difficulty slides contained four to six items, and weapons were less easily visible – either turned at a slight angle and/or partially obscured by other items. Challenging slides contained between five and nine items, and weapons were obscured and/or angled such that they were not easily discernable as weapons.

Some participants indicated that they were most likely to use the AWD on slides for which the correct answer was not immediately obvious. Therefore, it is likely that slide difficulty will be positively related to AWD use – in other words, participants will be more likely to use the AWD on more difficult slides than on easier slides. In contrast, because the presence or absence of weapons should be more obvious in easier slides than in more difficult slides, it is expected that participants will be more likely to correctly disagree with the AWD on easier slides rather than more difficult ones.

In order to test these post-hoc hypotheses, the proportions of use and correct disagreements made was calculated for easy, medium, and challenging slides. Dependent samples t-tests were employed to determine whether those proportions significantly differed from one another. In regard to automation use, the post-hoc hypotheses were supported. Results of the t-tests indicated that participants were significantly more likely to use the AWD on medium ($t=2.78$, $p=.01$) and challenging ($t=5.08$, $p<.01$) slides than

on easy slides. In addition, participants were significantly more likely to use the AWD on challenging slides than on medium slides ($t=2.65$, $p=.01$).

The hypotheses regarding the relationship of slide difficulty to automation monitoring were also supported. Dependent-samples t-tests indicated that participants were significantly more likely to correctly disagree with the AWD on easy ($t=15.48$, $p<.01$) and medium ($t=6.87$, $p<.01$) slides than on difficult slides. Participants were also significantly more likely to disagree with the AWD on easy slides than on medium slides ($t=9.34$, $p<.01$).

These post-hoc analyses revealed that slide difficulty played an important role in both automation use and automation monitoring. Participants were more likely to use the automated system on more difficult slides than on easier slides, and participants were more likely to correctly identify automation errors on easier slides than on more difficult slides. In regard to automated system use specifically, the effect of the trust and self-efficacy interaction decreased as the slide difficulty level increased. The relationships of affect and post-task trust showed the opposite pattern. As slide difficulty increased, the effects of affect and trust on automation use also exhibited a pattern of increasing correlations. The effects of faith and pre-test trust (both measured prior to the task) remained essentially constant across slide difficulty. It is also interesting to note that sex exhibited significant and increasingly negative relationships with automation use as slide difficulty increased such that males were less likely to use the automated system on more difficult slides than on easier slides.

Slide difficulty also exhibited interesting, although less consistent, relationships with automation monitoring. For both pre-task trust and the interaction of trust and self-

efficacy, the relationship with automation errors identified was near zero for easy and medium difficulty slides but became significant for difficult slides only. For both affect and post-task trust, the relationships exhibited a U-shape such that the correlations were significant for easy slides, decreased in magnitude (but were still significant) for medium slides, and reached their highest levels on difficult slides. The relationships found in these post-hoc slide difficulty analyses suggest that further research on interactions between slide difficulty and other constructs may yield interesting and valuable results.

Appendix P

Post-hoc Analyses of the Effects of Weapon Presence or Absence

The point system applied to the experimental task was designed such that participants could not achieve maximum scores by searching every bag or nearly every bag. In real life, however, the consequences of letting a weapon slip by undetected are much more severe than the consequences of taking the time to search many bags. The effect of the point system may not have been strong enough to overwhelm the powerfully engrained ideas about the relative consequences of under- and over-searching. It is possible that participants would have had a tendency to search whenever they believe that a weapon could possibly be present. Therefore, in regard to monitoring behavior, it was hypothesized that participants would be more effective in disagreeing with the automated system when it incorrectly suggests that a bag contains no weapon than when the machine incorrectly suggests that a weapon is present. Furthermore, it was hypothesized that this tendency would grow increasingly stronger with increasing slide difficulty, as the presence or absence of weapons becomes less obvious.

To examine this hypothesis, the proportion of correct disagreements was calculated separately for bags containing weapons and bags containing no weapons. Each of these proportions was calculated for each of the three slide difficulties, yielding six proportions.

First, dependent samples t-tests were employed to test for significant mean differences in correct disagreements between slides with no weapon and slides containing weapons within each difficulty level. A significant difference was found between slides containing weapons and slides not containing weapons for the easy difficulty slides (t = -

12.54, $p < .01$). The direction of the effect was as expected – participants were significantly more likely to disagree with an incorrect AWD recommendation when a weapon was present than when a weapon was not present. However, the effects for slides of medium difficulty (t = -.61, $p = .54$) and challenging difficulty (t = .99, $p = .32$) were both non-significant.

Next, t-tests were used to contrast the proportion of correct disagreements with AWD clears among the slide difficulty levels. In other words, differences were tested for slide difficulty on the likelihood that participants would search for a weapon when the AWD *incorrectly* suggested that none were present. Results indicated that participants were significantly more likely to correctly disagree with a suggested clear on easy slides than on medium slides (t = -15.22, $p < .01$) and on challenging slides (t = -19.75, $p < .01$). In addition, participants were more likely to correctly disagree with clears on medium slides than on challenging slides (t = -6.74, $p < .01$). The contrasts using correct disagreements with alerts yielded the same pattern. Participants were more likely to correctly disagree with incorrect alerts on easy slides than on medium slides (t = -3.09, $p < .01$) and on challenging slides (t = -9.91, $p < .01$). Participants were also more likely to correctly disagree on medium slides than on challenging slides (t = -6.57, $p < .01$). This structure of effects lends support to the slide difficulty categorizations and suggests that slide difficulty had similar effects regardless of weapon presence or absence.

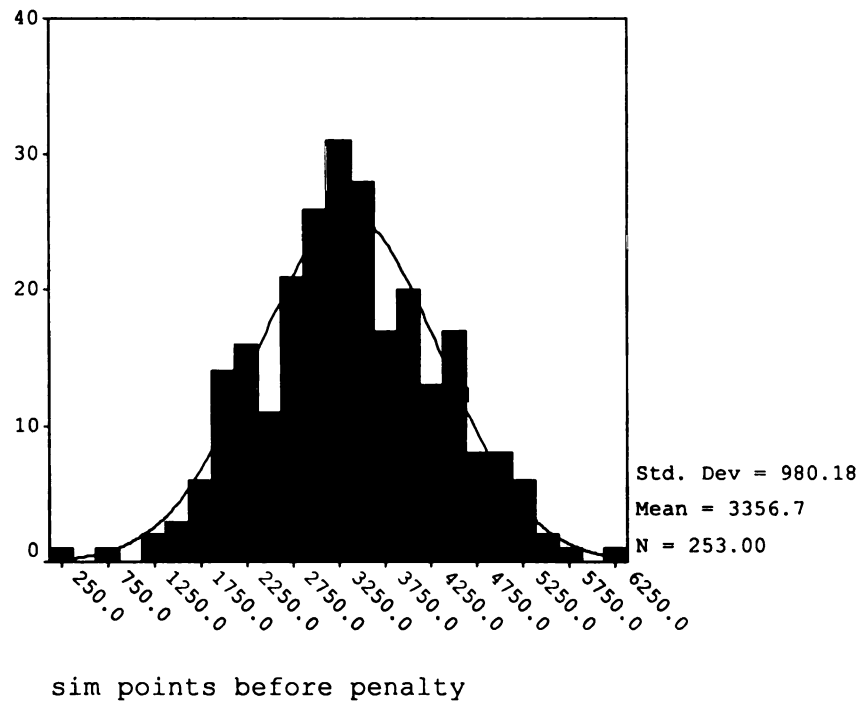
Appendix Q

Post-hoc Analyses of Point Performance

Task performance is a pervasive and important outcome in work-related studies. Therefore, several analyses were performed in relation to task performance, as measured by total points earned. A graph of the point totals reveals a relatively normal distribution (see Figure 12). However, two outliers were present. The following analyses exclude these two participants, who were removed from the analyses due to their extremely low point totals. For information on point means and standard deviations, see Table 21.

Figure 12

Distribution of Total Points



First, the effects of the three manipulations on total points were examined. A significant positive relationship was found between trust condition and total points such that those in the high trust condition (those with higher-functioning AWDs) scored more points than did participants in the low trust condition ($r = .29, p < .01$). However, no

significant relationships were found for self-efficacy condition (r = .02, p = .73) or affect condition (r = .03, p = .66).

Table 21

Descriptive Statistics for Total Points

	N	Min	Max	Mean	Std Dev
Points	253	250.00	6325.00	3356.72	980.18
AWDUse	253	00.00	1.00	.57	.44
Total Disagree	235	00.00	1.00	.18	.16

Next, correlations were examined between total points and the scales included in the study. No hypotheses were made regarding these relationships; they were examined from an exploratory standpoint. A significant negative relationship was found between automation-induced complacency potential and total points (r = -.13, p = .04). This relationship is interesting in that the relationship does not seem to operate through automation use or automation monitoring (as those correlations were non-significant). Instead, automation-induced complacency potential seems to operate on task performance through another mechanism not assessed in this study. Because of the attenuation resulting from the low internal consistency of this scale, it is likely that the true relationship between complacency potential and total points is actually stronger than the one found here. In addition, significant relationships with total points were found for post-task trust (r = .23, p < .01) and conscientiousness (r = -.15, p = .02). The relationships for extraversion (r = -.07, p = .25), desire to decrease cognitive load (r = -.07, p = .28), self-efficacy (r = -.03, p = .66), faith (r = -.04, p = .52), pre-task trust (r = .05, p = .47), and affect (r = .05, p = .39) were non-significant.

It was also hypothesized that total points could be predicted by the interaction of the trust condition and the extent to which the automated machine was used. It was predicted that participants in the high trust condition who used the machine more would have the most points and those in the low trust condition who used the machine more would have the least points.

Table 22

Regression Analysis of Trust Condition and AWD Use on Total Points

	Variable	N	β	R^2	ΔR^2
STEP 1	Trust Condition	129	432.17*	*.09*	
	AWD Use	129	-1563.29*		
STEP 2	Trust Cond	129	94.06		
	AWD USE	129	-6541.56*	*.14*	*.05**
	Trust Cond x USE	129	2795.20*		

*$p < .05$

As seen in Table 22, the interaction of trust condition and AWD use added significantly to the prediction of total points beyond the direct effects of those variables (R^2 change = .05, p = .01). As the graph in Figure 13 demonstrates, when AWD use was zero, trust condition had no effect. However, as AWD use increased, the participants in the low trust condition performed increasingly poorly relative to the participants in the high trust condition.

A similar proposition was tested using automation monitoring and trust condition (see Table 23). The interaction between trust condition and the proportion of correct disagreements produced a significant increase in prediction over and above the direct effects (R^2 change = .12, $p < .01$). A graph of this interaction reveals an interesting effect (see Figure 14). As expected, for participants in the low trust condition, points increased sharply as correct disagreements increased. However, for participants in the high trust

condition, points actually decreased somewhat as correct disagreements increased. This finding might be explained by the functioning of the AWD program. In the high trust condition, the AWD was programmed to provide a correct recommendation 85% of the time. However, the exact percentage will approach 85% over time; the machine is not correct exactly 85 out of 100 times. Perhaps the relatively brief length of this experimental task allowed for variance in the actual percentage of slides on which the AWD was correct. Therefore, some participants in the high trust condition may have had AWD machines with a greater likelihood of making mistakes. Because trust condition (and therefore AWD percentage correct) is significantly related to points, perhaps the decline in points as disagreements increase reflects an AWD that makes more mistakes and thus requires more disagreements.

Figure 13

Plot of the Interaction of Trust Condition and AWD Use onto Total Points

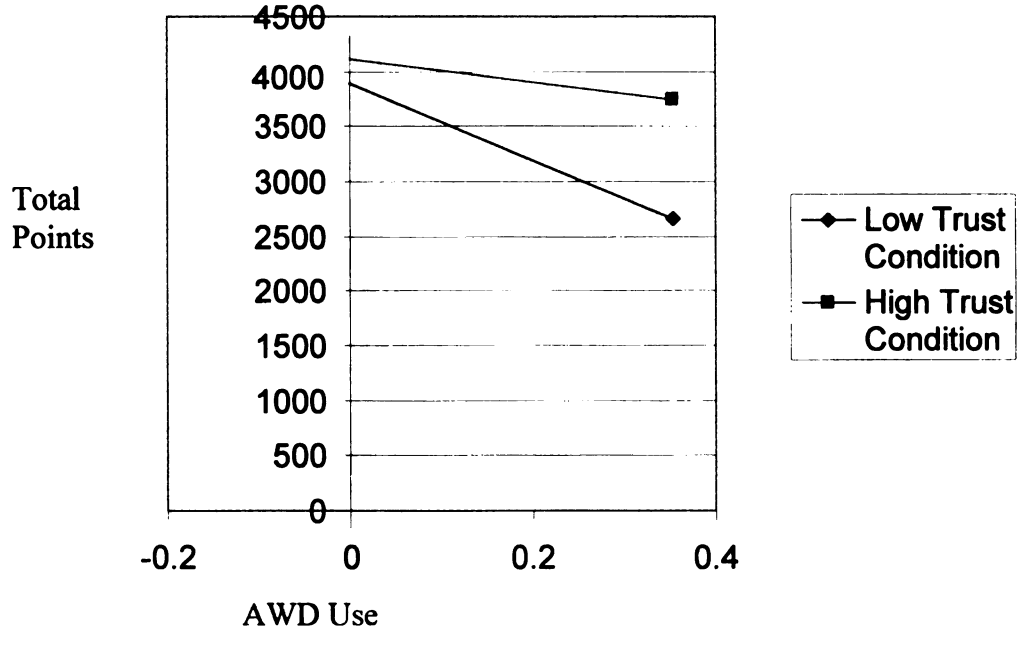


Table 23

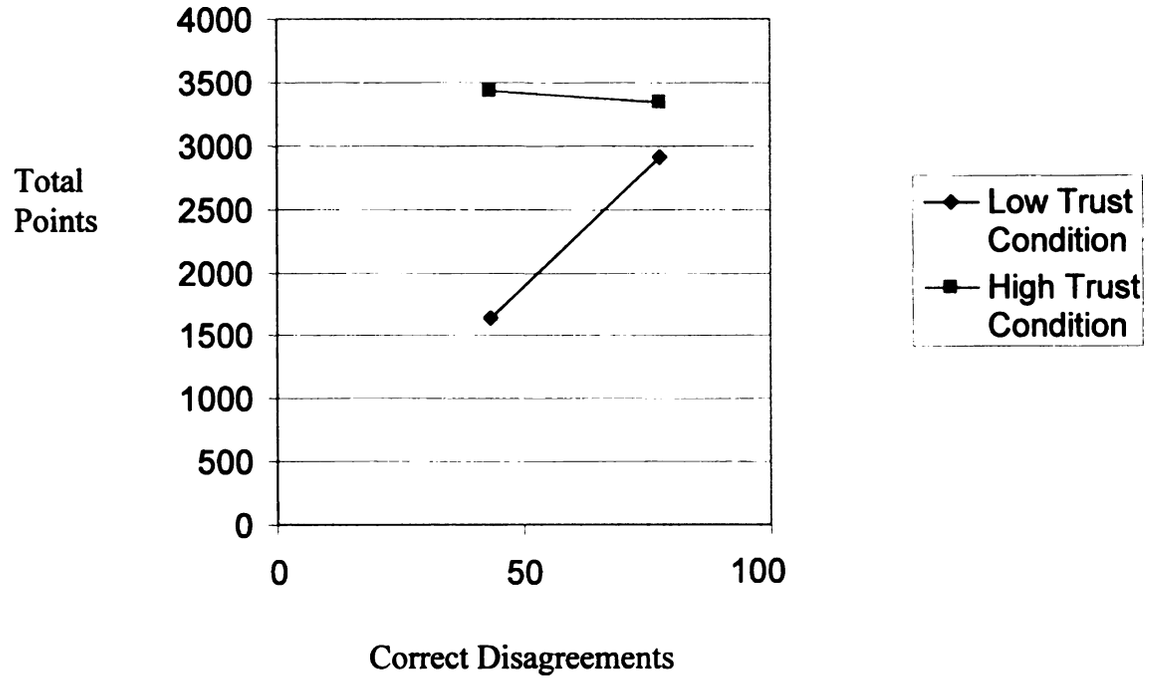
Regression Analysis of Trust Condition and Correct Disagreements on Total Points

	Variable	N	β	R^2	ΔR^2
STEP 1	Trust Condition	124	1049.21*	*.37*	
	Correct Disagreements	124	7.78*		
STEP 2	Trust Cond	124	3515.12*		
	Correct Disagreements	124	74.97*	*.51*	*.14**
	Trust Cond x Correct Disagree	124	-39.71*		

* $p < .05$

Figure 14

Plot of the Interaction of Trust Condition and Correct Disagreements onto Total Points



Finally, the effects of automation use and automation monitoring on total points were calculated. It is important to note that these results are sample specific and highly dependent on the level of functioning of the particular automated system in use here. The relationship between automation use and total points was tested using the data from session type A only. A significant and negative correlation was found between automated system use and total points ($r = -.23$, $p = .01$). This correlation indicates that

overall on this task, the more participants used the automated system, the lower their total point score. The relationships between monitoring and total points were tested using session type B only. It was interesting to find that the proportion of disagreements with the AWD related negatively to total points ($r = -.48, p < .01$), as did correct disagreements with the AWD on easy slides ($r = -.34, p < .01$). However, the relationships of total correct disagreements ($r = -.07, p = .45$), correct disagreements on medium slides ($r = .10, p = .28$), and correct disagreements on difficult slides ($r = -.13, p = .14$) with total points were non-significant. These negative correlations may reflect the trust condition – participants in the high trust condition needed fewer disagreements in order to get a higher score. This interpretation is supported by the strong correlation between trust condition and total points ($r = .59, p < .01$). In light of this fact, it is difficult to interpret these findings related to the relationships between automation use and automation monitoring.

Appendix R

Affect Manipulation Replication

The affect manipulation in the current study was included in an attempt to replicate the effect found by Reeves and Nass (1996). This effect was not replicated in the current study. There are at least two possible causes for this failure. First, the effect of machine name on affect toward the machine may be very small in size. The sample size of 253 in the current study may have been insufficiently large to detect a small effect. Second, the manipulation in the present study was present in a one-page set of instructions. Although the varied automated system name (either "the AWD" or "your automated assistant") appeared multiple times during the one page set of instructions, it is possible that one page was unable to saturate the participants with a manipulation strong enough for the effect to manifest. This hypothesis is supported by a non-significant correlation between affect condition and affect scores. In sum, it appears that either a very strong manipulation and/or a very large sample size is required in order to detect this particular manipulation.

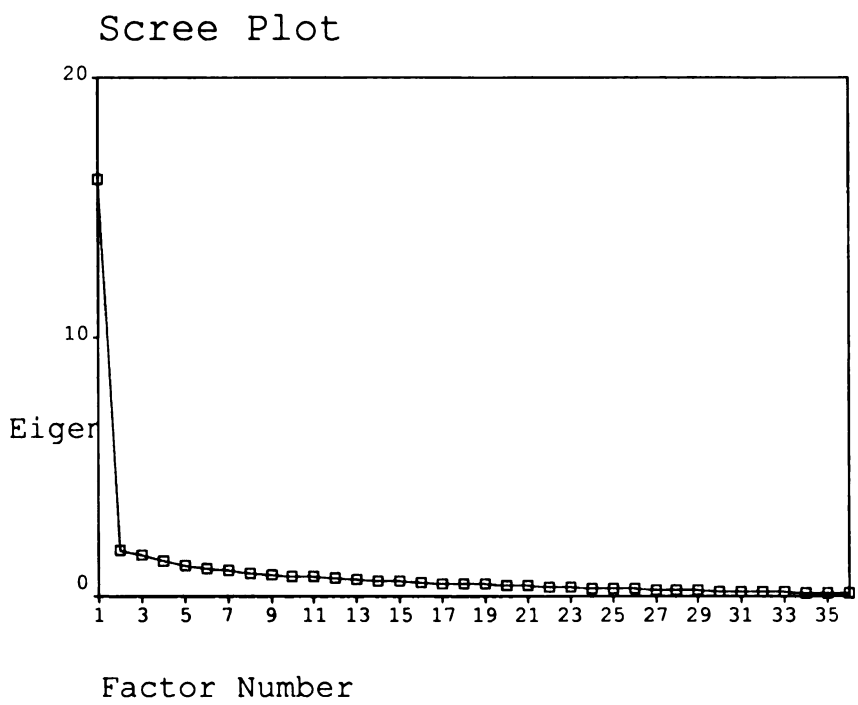
Appendix S

Trust Measure Analysis

Muir (1987) proposed that six trust factors would affect trust in an automated system. The 36-item post-task trust measure was composed of items representing five of these trust factors: predictability, dependability, faith, competence, and responsibility (recall that the sixth factor was measured as automation-induced complacency potential). A reliability analysis revealed that the 36 items had an alpha coefficient of .96, indicating a high degree of internal consistency. Furthermore, an exploratory factor analysis indicated that a strong initial factor on which 34 of the 36 items loaded. Figure 15 presents the scree plot for this factor analysis.

Figure 15

Scree Plot – Exploratory Factor Analysis of Trust Scale Items



Because competence, predictability, responsibility, and dependence were all manipulated together, firm conclusions regarding the divergent validity of these four constructs cannot be drawn from these results. On the other hand, faith and persistence (as automation-induced complacency potential) were not manipulated but measured. We should therefore be able to discriminate between these two scales. An exploratory factor analysis on the items of the faith and complacency potential scales do reveal the two expected factors – one factor on which all of the faith items load and one factor on which the majority of the complacency potential items load.