Old Dominion University
ODU Digital Commons

Psychology Theses & Dissertations | Psychology

Winter 2004

The Effects of Operator Trust, Complacency Potential, and Task Complexity on Monitoring a Highly Reliable Automated System

Nathan R. Bailey
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds

Part of the Industrial and Organizational Psychology Commons

Recommended Citation
Bailey, Nathan R.. "The Effects of Operator Trust, Complacency Potential, and Task Complexity on Monitoring a Highly Reliable Automated System" (2004). Doctor of Philosophy (PhD), dissertation, Psychology, Old Dominion University, DOI: 10.25777/wfmq-tv11
https://digitalcommons.odu.edu/psychology_etds/147

This Dissertation is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

المنارة للاستشارات

www.manaraa.com

THE EFFECTS OF OPERATOR TRUST, COMPLACENCY POTENTIAL, AND TASK COMPLEXITY ON MONITORING A HIGHLY RELIABLE AUTOMATED SYSTEM

by

Nathan R. Bailey
B.A. June 1998, Ohio University
M.S. August, 2001, Old Dominion University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2004

Approved by:

Mark W. Scerbo (Director)

James P. Bliss (Member)

Peter J. Mikulka (Member)

Prinzel (Member)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

THE EFFECTS OF OPERATOR TRUST, COMPLACENCY POTENTIAL, AND TASK COMPLEXITY ON MONITORING A HIGHLY RELIABLE AUTOMATED SYSTEM

Nathan R. Bailey
Old Dominion University
Director: Mark W. Scerbo

Technological advances have allowed for widespread implementation of automation in complex systems. However, the increase in quantity and complexity of advanced automated systems has raised a number of potential concerns including degraded monitoring skills. The present investigation consisted of two studies that assessed the impact of system reliability, complacency potential, monitoring complexity, operator trust, and system experience on monitoring performance. In both studies, participants monitored a simulated aviation display for failures while operating a manually controlled flight task. In addition, the second experiment assessed the ability of operators to detect a single automation failure over three experimental sessions. Results indicated that realistic levels of system reliability severely impaired an operator's ability to monitor effectively. In addition, as system experience increased, operator performance for monitoring highly reliable systems continued to decline. Further, operators who reported higher levels of trust, confidence, and more frequent usage of automation demonstrated poorer overall monitoring. The complexity of the monitoring task was also shown to be one of the most important factors influencing operator monitoring performance with poorer performance on more cognitively demanding tasks that continued to degrade as system experience increased. Results from both studies indicated that operator trust increased as a function

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of increasing system reliability and that as trust increased, monitoring performance decreased. These results suggest that for highly reliable systems, increasing task complexity and extensive experience may severely impair an operator's ability to monitor for unanticipated system states.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGMENTS

I would first like to thank my academic advisor, Dr. Mark Scerbo, for his input on this document and for his support throughout the tenure of my graduate experience. You have been a true mentor. I would also like to thank my committee members including Dr. James Bliss, Dr. Peter Mikulka, and Dr. Lawrence Prinzel III for their guidance and support as well as Dr. J. Raymond Comstock for his help developing the experimental tasks used in this research. In addition, I would like to acknowledge my friends and family for their encouragement throughout the years. Finally, I would like to thank Hope. Your love and support has been invaluable.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Figures Page

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

INTRODUCTION

Automation can be characterized as the execution by a machine of a function that was previously carried out by a human (Parasuraman & Riley, 1997). The widespread implementation of automation in complex systems such as transportation, maintenance, process control, decision support systems, and quality control has been the result of anticipated improvements in system performance, efficiency, and safety. These improvements have been generally realized. Within the context of commercial aviation, automated systems have made it possible to reduce flight times, improve fuel efficiency and passenger comfort, navigate more effectively, and improve the perceptual and cognitive abilities of crewmembers (Wiener, 1988). However, the increase in quantity and complexity of advanced automated systems has raised a number of real and potential concerns including increased operator workload, loss of task proficiency, reduced situation awareness, and degraded monitoring skills (Endsley, 1996; Parasuraman, Molloy, & Singh, 1993; Wiener & Curry, 1980; Wiener, 1988).

With the increase in quantity and complexity of advanced automated systems has come an increased demand for operators to monitor systems for failures or unanticipated states (Sarter & Woods, 1995; Wiener & Curry, 1980). One negative consequence that may result from increased monitoring demands has been referred to as automation-induced complacency (Parasuraman et al., 1993; Wiener, 1981). Automation-induced complacency is thought to exist in highly reliable automated systems where an operator serves in a backup role and refers to the decline in monitoring performance that often follows the shift from performing a task manually to monitoring the automation of that task (Farrell & Lewandowsky, 2000; Parasuraman et al., 1993). The following account

The journal model for this dissertation is *Journal of Experimental Psychology: Applied.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

taken from NTSB report 1-0016 is one of the first aviation accidents attributed to automation-induced complacency.

Eastern Airlines Flight 401

On December 29, 1972, Eastern Air Lines flight 401 (EAL 401) departed from John F. Kennedy International Airport (JFK), Jamaica, New York at 2120 EST bound for Miami International Airport (MIA), Miami, Florida. The Lockheed L-1011 was carrying 143 passengers and 13 crew members. The flight was uneventful until their approach into MIA where they encountered a possible problem with their front landing gear. After the flightcrew lowered the landing gear, a green light indicating that the front gear was firmly locked into place failed to illuminate. Subsequently, the captain recycled the gear but the indicator still did not light. The following transcription is the result of data taken from the digital flight data recorder system (DFDR) and the cockpit voice recorder (CVR) on EAL 401.

2334:05 - EAL 401 called the MIA tower and stated, "Ah, tower this is Eastern, Ah, four zero one, it looks like we're gonna have to circle: we don't have a light on our nose gear yet."

2334:14 - The tower advised, "Eastern four oh one heavy, roger, pull up, climb straight ahead to two thousand, go back to approach control, one twenty eight six."

2335:09 - EAL: 401 contacted MIA approach control and reported, "All right, ah, approach control, Eastern four zero one, we're right over the airport here and climbing to two thousand feet, in fact, we've just reached two thousand feet and we've got to get a green light on our nose gear."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2336:04 - The captain instructed the first officer, who was flying the aircraft, to engage the autopilot. The first officer acknowledged the instruction.

2336:27 - MIA approach control requested, "Eastern four oh one, turn left heading three zero zero." EAL 401 acknowledged the request and complied.

2337:08 – The captain instructed the second officer to enter the forward electronics bay, below the flight deck, to visually check the alignment of the nose gear. This check involved viewing the physical alignment of two rods on the landing gear linkage which could be seen through an optical sight located in the forward electronics bay.

2337:24 – A downward vertical acceleration transient of 0.04 g caused the aircraft to descend 100 feet; the loss in altitude was arrested by a pitchup input.

Meanwhile, the flightcrew continued their attempts to free the nose gear position light lens from its retainer, without success. At 2338:34, the captain again directed the second officer to descend into the forward electronics bay and check the alignment of the nose gear indices.

2338:56 until 2341:05, the captain and the first officer discussed the faulty nose gear position light lens assembly and how it might have been reinserted incorrectly.

2340:38 – A half-second C-chord, which indicated a deviation of ± 250 feet from the selected altitude, sounded in the cockpit. No crewmember commented on the C-chord. No pitch change to correct for the loss of altitude was recorded.

2341:40 - MIA approach control asked, "Eastern, ah, four oh one, how are things comin' along out there?" This query was made a few seconds after the MIA controller noted an altitude reading of 900 feet in the EAL 401 alphanumeric data block on his radar display. The controller testified that he contacted EAL 401 because the flight was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

nearing the airspace boundary within his jurisdiction. He further stated that he had no doubt at that moment about the safety of the aircraft. Momentary deviations in altitude information on the radar display, he said, are not uncommon; and more than one scan on the display would be required to verify a deviation requiring controller action.

2341:44 - EAL 401 replied to the controller's query with, "Okay, we'd like to turn around and come, come back in" and at 2341:47 approach control granted the request with; "Eastern four oh one turn left heading one eight zero." EAL 401 acknowledged and started the turn.

2342:05 - The first officer said, "We did something to the altitude." The captain's reply was, "What?"

2342:07 - The first officer asked, "We're still at two thousand, right?" and the captain immediately exclaimed, "Hey, what's happening here?"

2342:10 - The first of six radio altimeter warning "beep" sounds began; they ceased immediately before the sound of the initial ground impact.

2342:10 - While the aircraft was in a left bank of 28°, it crashed into the Everglades 18.7 miles west-northwest of MIA. The aircraft was destroyed by the impact.

The crash killed 96 passengers and 5 crew members. After examination of the nose gear warning light, it was determined that both bulbs in the unit had burned out. It was further confirmed that the front gear had, in fact, been locked into position. As concluded by the investigating committee, the force applied to the control column at 2337:24 was sufficient to disengage the altitude hold automation mode. The most likely cause of the force applied to the control column was inadvertent contact by either the captain or the first officer while moving around the cockpit. Although such an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

occurrence should have been evident with the extinguishing of the altitude mode select light on the annunciator panel, it was later found that as a result of a miscalibration between the captain's controls and those of the first officer, it was possible that only the annunciator on the captain's side of the aircraft would have indicated the disengagement. In conjunction with the unintended mode change, a number of reductions in power were also made by the flightcrew to compensate for excess airspeed. The altitude hold disengagement in combination with the subsequent decreases in engine power resulted in the uncommanded descent and eventual crash of the aircraft (NTSB, 1973).

The probable cause of the accident was determined to be the failure of the flightcrew to monitor flight instrumentation during the final minutes of flight and to detect the unexpected descent quickly enough to prevent the crash. Preoccupation with the malfunction of the nose landing gear indicator distracted the crew's attention from the flight instruments which allowed the unintended descent to go unnoticed. However, according to the investigating committee, regardless of the manner in which the autoflight system status was represented to the crew, the flight instruments, (e.g., altimeters, vertical speed indicators, airspeed indicators, pitch attitude indicators, and the autopilot vertical speed selector), would have indicated nonlevel flight conditions. Taken together, the altitude-alerting C-chord signal and the flight instrument indications should have alerted the crew to the undesired descent. Members of the committee further emphasized their concerns with the new automated systems which were becoming widely used at the time. They argued that flightcrews were growing steadily more reliant on the functioning of aviation automation, especially as its reliability increased. As a result, manual operations, basic supervision, and monitoring of flight status by the instrumental

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

indicators would suffer. The crew's overreliance on automated systems and the resulting decline in monitoring performance that led to the crash highlights one of the potential dangers associated with highly automated systems.

Automation-Induced Complacency

Although the concept of automation-induced complacency has been discussed for many years, an acceptable definition has been difficult to generate. Billings, Lauber, Funkhouser, Lyman, and Huff (1976) defined automation-induced complacency as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state" (p. 23). Wiener (1981) defined automation-induced complacency as a "psychological state characterized by a low index of suspicion" (p. 117). Farrell and Lewandowsky (2000) offer a definition that relies more on the relationship between manual control and monitoring, suggesting that complacency refers to the ensuing decline in performance that occurs when individuals shift from performing a task themselves to monitoring its automation.

Despite the lack of consensus among definitions, complacency has long been implicated in aviation accidents (Hurst & Hurst, 1982; NTSB, 1973; Wiener & Curry, 1980) with two key factors present in most cases. First, operators tend to be less aware of system states when automation is performing a function for them, especially if they are simultaneously engaged in other tasks. Second, operators of complex systems are not well suited for monitoring infrequent and unexpected problems, especially in highly reliable systems (Wiener & Curry, 1980). The crash of EAL flight 401 provides a tragic but cogent example of automation-induced complacency that resulted from the crew's preoccupation with the landing gear indicator malfunction. The crew's focus on that task

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

resulted in their failure to detect the unintended disengagement of the altitude hold automation, despite multiple instrumental readings and an auditory warning that should have alerted them to the deviation. Their focus on the malfunctioning indicator (i.e., doing another task) in conjunction with the inadvertent disengagement of the altitude hold (i.e., an infrequent and unexpected problem), led to the eventual crash of the aircraft. However, despite the general acknowledgement that automation-induced complacency could negatively impact human performance and aviation safety, little effort was aimed at describing the construct and its underlying mechanisms (Wiener, 1981).

Empirical Research on Automation-induced Complacency

In response to Earl Wiener's (1981) criticism that complacency was largely an anecdotal construct, the first empirical study of automation-induced complacency was conducted by Thackray and Touchstone (1989) using an air traffic control (ATC) simulation. In their study, 40 participants were divided into two experimental conditions. The first included automation that aided participants in the detection of critical incidents. In the second condition, participants received no automated aiding. For the critical incidents, two different stimuli were used including a simple and complex monitoring task. The simple monitoring task consisted of detecting a series of *X*'s that had replaced an aircraft's altitude reading. The more complex monitoring task required participants to integrate heading and altitude information and determine if two aircraft were in a potential conflict. The first critical event was considered readily detectable and was described as a malfunction of the aircraft's transponder. The second critical event was regarded as substantially more difficult since it was not immediately apparent and required a check of multiple parameters.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Each participant completed four 30-min trials with nine critical incidents in each trial including three X's, three nonconflicting altitude changes, and three conflicting altitude changes. In each case, participants were required to press a key signaling their detection of either the automated aid or the actual incident. In the case of detecting the X's, no further input was required. However, in the case of detecting potential flight path conflicts, participants were required to respond to both the detection of the potential conflict and to provide a valid change in altitude to avoid the collision.

The same display was used in each experimental condition with the exception that half of the participants received advisory alerts regarding potential malfunctions and conflicts. However, these advisories were programmed as if they failed to detect conflicting aircraft on two separate occasions with failures limited to only the potential conflicts category of incidents. One of the automation failures occurred in the first trial of the experiment during the first half hour and the second failure occurred in the final 15 min of the 2-hr experimental session.

Thackray and Touchstone (1989) hypothesized that participants who received automated aiding would become increasingly dependent on the aid and would reduce their efforts to monitor potential conflicts. Thus, detection response times and miss rates for conflicts the automation failed to detect would exceed those where no automated aid was given. It was further hypothesized that monitoring efficiency for those in the advisory alert condition would suffer more in the latter portion of the experiment. However, Thackray and Touchstone found that participants detected the potential ATC conflicts equally well in both conditions. Response times for detecting the simple alphanumeric change in the transponder malfunctions revealed no significant increase

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

over the course of the experiment. There were also no significant differences between the detection rates for participants with and without the automated aid. Finally, the authors failed to demonstrate a difference in detection times in conflict monitoring for those conflicts that occurred early versus late in the experimental session.

Clearly, this study provides limited empirical evidence for automation-induced complacency. However, Thackray and Touchstone (1989) indicated that their failure to obtain compelling empirical evidence may have been the result of a relatively short experimental session, stating:

> Although studies such as this are of value in helping to define those parameters that may or may not contribute to the development of complacency effects, definitive answers to the difficult questions posed above may well require lengthy field studies in which infrequent errors or failures are introduced while performing under real-life or highly realistic simulated conditions. (p. 9)

Thackray and Touchstone's study has also been criticized by Parasuraman et al. (1993) on the grounds that participants only operated a single task. Parasuraman et al. argued that any performance consequences for automation-induced complacency were more likely to exist in environments where operators had multiple concurrent duties and were responsible for more than just a simple monitoring task.

Accordingly, Parasuraman et al. (1993) conducted a set of studies to determine if they could find performance effects related to complacency where Thackray and Touchstone (1989) had failed. Parasuraman et al. had four hypotheses. First, they argued complacency would be high for a group of participants who encountered automation with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

constant, unchanging reliability. In contrast, participants who experienced automation with variable reliability would be less likely to exhibit complacency. Second, the authors believed that the initial level of reliability was important and that those participants encountering higher initial levels of reliability would have a greater potential for complacency. Third, because trust in automation is generally reduced immediately following a failure (Lee & Moray, 1992), several consecutive failures should reduce the effects of complacency with a corresponding increase in monitoring performance. Finally, given the inability of Thackray and Touchstone to demonstrate any performance consequences with a single task, Parasuraman et al. proposed that all predictions would hold only when operators were responsible for completing multiple concurrent tasks.

In their first experiment, Parasuraman et al. (1993) had 24 participants operate a modified version of the Multi-Attribute Task Battery (MAT; Comstock & Arnegard, 1992). The MAT is a suite of flight simulation tasks including compensatory tracking, resource management, system monitoring, communications, and scheduling; however, Parasuraman et al. used only the compensatory tracking, resource management, and system monitoring portions of the MAT. The compensatory tracking task requires participants to use a joystick to maintain the position of a constantly moving circle as close to the center of a target as possible. The resource management task requires participants to maintain a constant level of fuel in two primary tanks by moving fuel from other tanks using a series of pumps. The system monitoring task requires participants to detect deviations from a center value on four vertical gauges that represent the temperature and pressure of two engines. Under normal conditions, malfunctions in the monitoring task were detected automatically and participants were not required to make

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

any corrections. However, the reliability of the automated system for the monitoring task was varied and not all deviations were detected. These automation failures required participants to make a keyboard input to bring the system back to a normal state. In a second experiment, participants were required to perform only the system monitoring portion of the MAT.

Parasuraman et al. (1993) found that automation complacency effects were eliminated when the reliability of the automated system was variable, alternating between high and low, with improved monitoring performance for those participants under variable reliability. Their second hypothesis regarding the initial level of system reliability was not supported. The performance of those participants who experienced higher initial levels of automation reliability did not differ from participants whose initial level of automation reliability was lower. The authors also found only partial support for their hypothesis that following a number of consecutive failures, monitoring performance would increase. Although monitoring performance did increase after a number of failures, it did not achieve the same level associated with the variable reliability condition. Finally, by comparing their first experiment to the second, Parasuraman et al. demonstrated that the performance consequences of complacency were limited to conditions that required operation of multiple concurrent tasks. These findings illustrate some of the first empirical performance implications regarding automation-induced complacency.

Molloy and Parasuraman (1996) conducted a follow-up study that examined task complexity and the effects of time on monitoring for a *single* automated failure. Their experiment employed a modified version of the MAT and used three different levels of task complexity. In the multi-complex condition, participants were responsible for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

performing the compensatory tracking, system monitoring, and resource management portions of the MAT. In the single-complex condition, participants were required to perform only the system monitoring task. The third level of task complexity consisted of a simple visual task that required participants to detect a nonstandard stimulus over successive presentations. Although both the single-complex and the simple visual tasks each required operators to detect single discrete events, the simple visual task was regarded as significantly less demanding. Molloy and Parasuraman predicted that individuals in the multi-complex task would be less likely to detect the single failure because their attention would be divided among multiple concurrent tasks. Individuals in both the multi-complex and simple visual task conditions were also expected to exhibit better detection performance at the beginning than at the end of each session. This expectation is consistent with findings that performance can become degraded in settings where operators monitor systems with very low signal rates and acknowledges the impact that dividing attention among multiple tasks over an extended period may have on detection performance (Loeb & Binford, 1970). Finally, participants in the single-complex task condition were expected to demonstrate improved detection performance due to the increased attentional resources resulting from their limited task responsibilities.

As expected, participants in the multi-complex condition demonstrated degraded monitoring efficiency for detecting the single automation failure. Molloy and Parasuraman (1996) also found that monitoring performance degraded over time for both the multi-complex condition and the simple visual task. Finally, those participants in the single-complex task, whose responsibilities were limited only to the system monitoring task, demonstrated highly accurate monitoring for the single failure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Molloy and Parasuraman's (1996) results are important because they demonstrate the effects of task complexity on monitoring performance and do so in an environment that included a more realistic proportion of overall system failures. Further, the results extend the findings of Parasuraman et al. (1993), demonstrating that human monitoring of automation is inefficient for detecting single, infrequent failures which are more likely in highly reliable systems. The findings also bolster the assertion that highly reliable systems can engender poor monitoring performance as a result of overreliance or excessive trust in automated devices (Muir, 1989; Parasuraman et al., 1993).

Automation Reliability and Consistency

Previous research has shown that the reliability of an automated system impacts an operator's ability to monitor that system (Lee & Moray, 1992; Muir & Moray, 1996; Parasuraman et al., 1993). Lee and Moray demonstrated that both trust and strategies for using automation varied according to its overall reliability. Specifically, highly reliable systems induce trust, which impacts an operator's reliance on automation. Although the issue of trust in automation will be discussed in more depth in a subsequent section, the results of Lee and Moray suggest that operators are less likely to monitor highly reliable systems. This view is also consistent with the observations of Parasuraman et al. who found that when individuals operated highly reliable and consistent automated devices, they had poorer monitoring performance. By contrast, if the automated system exhibited lower and inconsistent levels of reliability, better overall monitoring performance was achieved. Muir (1987, 1994) has also argued that increasing system experience in highly reliable settings will further degrade monitoring performance as system experience accumulates.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Despite the evidence that system reliability is a fundamental factor impacting monitoring performance, the exaggerated proportions of system failure used in previous studies on automation-induced complacency make it difficult to draw conclusions regarding the impact of reliability on monitoring. In fact, Parasuraman et al. (1993) express criticism in their use of artificially high proportions of system failure that would be unacceptable in any real-world setting. They suggest further that there is a need to conduct research on automation-induced complacency using levels of reliability that approach or exceed 99%, over a number of experimental trials. Given that the majority of empirical research on complacency has used rather high proportions of system failure, it is reasonable to assume that the development of trust described by Muir (1987) and Rempel, Holmes, and Zanna (1985) may be stunted, yielding qualitative differences in how operators interact with and monitor the system. The elevated proportions of system failures typically cited as eliciting complacency may not establish any absolute sense of trust because individuals are invariably skeptical of system performance. Interacting with more realistic, highly reliable systems may in fact be considerably different from interacting with systems that exhibit only moderate levels of reliability. It is therefore necessary to elaborate on the findings of Parasuraman et al. and Molloy and Parasuraman (1996), incorporating a more realistic proportion of system failures, in conjunction with a longer experimental timeframe. These methodological changes will help to elucidate the impact that extensive system experience and more realistic levels of system reliability have on automation-induced complacency.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Impact of High Reliability on Operator Attentional Resources

One way that system reliability can impact monitoring performance is by affecting an operator's attentional resources. The resource theory of attention described by Kahneman (1973) considers the attentional resources of operators to be finite and that an operator's available resources are directly proportional to his or her level of arousal. Kahneman suggested that mental workload could be described as the discrepancy between task demands and an operator's available attentional resources. He went on to argue that only a certain number of tasks at a certain level of difficulty could be successfully completed before individuals began experiencing increased workload and/or degraded performance. By contrast, Young and Stanton (2002) have recently proposed a theory suggesting that operator "underload", (i.e., periods where workload is very low), is also related to decreased attentional capacity and degraded operator performance.

Malleable Attentional Resources Theory (MART) posits that during times of low workload, the attentional capacity of operators shrinks in much the same way it is exhausted when task demands are high. By examining operator performance and mental workload for driving tasks that used different forms of automation, Young and Stanton found that attentional capacity was positively related to mental workload. They argued that as workload decreased, so did the attentional capacity of operators. This decrease in attentional capacity may have important implications for monitoring performance in complex systems. Specifically, highly reliable systems may elicit lower levels of workload because they demand limited effort on the part of the operator to monitor for failures or unanticipated states. Consequently, as individuals operate these systems over extended periods, their attentional resources are further abated making it difficult to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

detect critical deviations on the rare occasions when they do occur. The decreased levels of arousal/workload associated with highly reliable systems may therefore help to explain the degraded monitoring performance associated with automation-induced complacency.

Reliability's Impact on Monitoring Performance for Unrelated Systems

Another important issue with respect to system reliability is whether the reliability of one system impacts operator monitoring performance on another unrelated system. Research by Muir and Moray (1996) found that distrust in one function of an automated system could spread and create distrust in another automated function controlled by the same component. This effect, however, was limited to components controlled by the same unreliable automated device. Muir and Moray did not find any generalization of distrust to other independent components in the same system or to entirely separate systems. However, as noted, the artificially high proportion of system failures used in previous research may have impacted the development of operator trust (Lee & Moray, 1992; Muir and Moray, 1996). Because automation-induced complacency is characterized by an often subtle but distinct loss of operator engagement, highly reliable systems have the potential to elicit this effect and are more likely to degrade monitoring performance on other unrelated tasks. By contrast, previous research may have yielded qualitatively different levels of operator trust and monitoring performance because the experimental tasks used were sufficiently engaging based on the need for operators to constantly monitor a system that was likely to fail. As such, it is important to investigate further the effects of reliability to determine whether a high degree of reliability will affect an operator's ability to monitor effectively for critical deviations in an unrelated system.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Within the present investigation, system reliability was one of the primary experimental manipulations. Specifically, operators working with a highly reliable system were expected to demonstrate degraded monitoring performance with respect to both detection rate and response time. This view is consistent with the findings of Lee and Moray (1992) and Parasuraman et al. (1993) as well as the degraded attentional capacity described in Young and Stanton's (2002) MART. Accordingly, the present investigation included two studies. The first used a level of system reliability that approximated 98.0%, as suggested by Parasuraman et al., within the high reliability condition. In addition, the low reliability condition utilized a level of system reliability that approached the high reliability condition from Parasuraman et al., (i.e., 87.0%). In the second study, an operator's ability to detect a single failure over multiple sessions was examined, with the reliability of the automated systems exceeding 99.7%. Using more realistic levels of system reliability addressed one of the primary criticisms of previous research on automation-induced complacency. In addition, these levels of reliability allowed for a more direct comparison between the present study and the research by Parasuraman et al. and Molloy and Parasuraman (1996). It was anticipated that individuals operating under high reliability would demonstrate degraded monitoring performance relative to individuals operating a lower reliability system.

Another criticism of previous research on automation-induced complacency is that the durations used were fairly brief. A number of researchers, including Lee and Moray (1992) and Muir (1987, 1994) have argued that as one's system experience increases, there is a qualitative shift in how one interacts with and monitors a system. Further, Muir has suggested that in high reliability systems, operator monitoring

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

performance will continue to degrade as experience with the system increases, (i.e., a negative relationship between system experience and monitoring performance). It was therefore important to examine the impact of increasing system experience on monitoring performance. Within the present study, individuals operating under high reliability were expected to demonstrate poorer monitoring performance across trials. Additionally, the present investigation examined the effect of system reliability on monitoring another unrelated system. It was expected that individuals in the high reliability condition would detect fewer failures in an unrelated monitoring task.

Complacency Potential

Related to system reliability and the impact it has on operator trust, individuals may also exhibit relatively persistent attitudes regarding technology that contribute to the style and effectiveness of their interaction with automated systems. An attitude has been defined as a personal disposition common to individuals but possessed in varying degrees, compelling them to react to objects and situations in favorable and unfavorable ways (Ajzen & Fishbein, 1980). As such, individuals may bring with them preexisting notions regarding automated devices that will influence the overall style, appropriateness, and efficiency of their interactions. These attitudes may increase or decrease the potential for automation-induced complacency.

Singh, Molloy and Parasuraman (1993) have argued that the potential for automation-induced complacency must be differentiated from those behaviors associated with complacency and that these attitudes may be related to Langer's (1989) concept of premature cognitive commitment. Premature cognitive commitment develops when an individual is initially exposed to a stimulus, device, or event within some specific context.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

That individual's initial attitude is then reinforced when he or she encounters the same stimulus within the same context. Langer argues that repetition of experience is one of the main antecedent conditions for premature cognitive commitment. Therefore, for operators who experience high reliability during their initial encounter with a system, each subsequent encounter where the system exhibits high reliability will reinforce their preexisting attitude. The concept of premature cognitive commitment is related to the confirmation bias whereby individuals tend to seek information that confirms a previous hypothesis and ignore information that is inconsistent (Fischoff & Beyth-Marom, 1983; Klayman & Ha, 1987). Therefore, in the case of highly reliable systems, operator attitudes will become more complacent over time as a result of their initial experience and the subsequent reinforcement of that experience over time.

The Complacency-Potential Rating Scale

In an attempt to determine whether the potential for complacency could be measured, Singh et al. (1993) developed a 20-item instrument, the Complacency-Potential Rating Scale (CPRS), that measures attitudes toward common automated devices. Singh et al. argued that complacent behaviors may manifest themselves when complacency potential exists in conjunction with a specific set of conditions including pilot inexperience with equipment or situations, excessive workload, fatigue due to poor sleep or long flights, and inefficient communication between crew members or between crew members and ground support. Complacency potential, therefore, represents a maladaptive attitude toward automation that may arise in certain contexts and adversely impacts operator performance.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Singh et al. (1993) were able to demonstrate that attitudes toward automation could be reliably measured and a number of other researchers have found utility for using the CPRS as a predictor of monitoring performance (see Bailey, Scerbo, Freeman, Mikulka, & Scott, 2003; Prinzel, DeVries, Freeman, & Mikulka, 2001). Therefore, it appears that attitudes toward automation in and of themselves may not significantly influence operator monitoring behavior. However, given the existence of certain circumstances, preexisting attitudes may play an important role in determining the appropriateness and efficiency of human-automation interaction.

Complacency Potential and Cognitive Task Demands

Although technology-related attitudes may by themselves influence operator performance, the cognitive demands of the task may further degrade an operator's ability to monitor effectively, especially if he or she has high complacency potential. Operators with high complacency potential are more likely to possess degraded attentional resources. As Young and Stanton (2002) indicate, operators who tend to allow automated systems to complete their responsibilities with little monitoring/intervention experience reduced task demands. However, according to MART as task demands are reduced, so too are the attentional resources available for completing subsequent tasks. High complacency potential operators will, therefore, experience degraded attentional resources owing to the reduced demands of the task that result from their suboptimal attitudes toward automated systems. Consequently, performance on more difficult and cognitively demanding tasks will be poorer for operators with high complacency potential because of their already degraded attentional resources.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Within the present set of studies, individuals who were high in complacency potential were expected to exhibit degraded monitoring performance. In addition, it was expected that individuals operating under higher levels of reliability, who possessed higher complacency potential, would be less able to monitor effectively. Poor monitoring performance was expected because their initial experience with the system would establish an attitude that the system was highly reliable. This attitude would then be reinforced over the duration of the experimental trials. It was also anticipated that individuals high in complacency potential would show greater deficiencies in monitoring performance over time due to their already complacent disposition and the repetitive nature of the experimental task. Finally, those individuals who were high in complacency potential were expected to have greater difficulty detecting system failures in a more cognitively demanding monitoring task due to the cognitively demanding nature of the task and the operator's predisposition toward complacent behavior.

Complexity of the Monitoring Task

Although system reliability and operator attitudes may be instrumental for eliciting automation-induced complacency, the intrinsic properties of the monitoring task may also influence monitoring performance. Both the degree of complexity and the cognitive resources required to adequately perform monitoring tasks may be important factors influencing human-automation interaction. However, previous research on automation-induced complacency has limited operator monitoring responsibilities to detecting simple discrete events, (e.g., an engine indicator exceeding some prespecified parameter). Aside from research by Thackray and Touchstone (1989), which did require operators to monitor for multiple types of failures that varied in difficulty, other empirical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

research on complacency has been limited to malfunctions that occur in only one portion of the interface, requiring very few cognitive resources besides the perceptual ability to discriminate a signal. Because operators of automated systems are often required to detect complex patterns composed of events that take place in divided portions of an interface, it is important to examine how the complexity of a task influences monitoring performance. As such, Grubb, Warm, Dember, and Berch (1995) conducted a study examining the effects of multiple-signal discrimination on vigilance performance and workload for complex displays. Specifically, they used a display that required operators to monitor 1, 2 or 4 portions of an interface for different critical signals. They found that as the number of displays that needed to be monitored increased, the ability of operators to correctly detect signals decreased. Therefore, as the attentional demands of the monitoring task went up, monitoring performance became degraded. They also discovered a positive relationship between perceived workload and the overall number of displays monitored. These results demonstrate the impact that more difficult monitoring activities have on the availability of attentional resources and operator workload.

In contrast with previous research on automation-induced complacency, the present investigation used a cognitively demanding monitoring task. First, consistent with Grubb et al. (1995), operators were asked to monitor multiple systems for different forms of critical deviations. In addition, one of the monitoring tasks required operators to memorize both the normal operating range for several engine parameters as well as the appropriate corresponding response for each parameter if a critical deviation was detected. The three monitoring tasks in conjunction with the operation of a primary flight task

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

demanded a sufficient proportion of operator resources to allow for a more ecologically valid examination of monitoring performance in complex environments.

The Pattern of Failures

In addition to examining the performance implications of operators monitoring multiple systems and a more cognitively demanding task, the pattern of failures for the more difficult task was also manipulated. As noted, operators often experience complex patterns of system failure. Because one of the overall goals of the present study was to utilize an ecologically valid setting for examining complacency, it was important to use a plausible pattern of automation failures. In most complex systems, failures are not randomly distributed. Instead, failing components or processes tend to break down and impact the reliability and functioning of related and/or subordinate systems. This pattern of failures may ultimately impact both operator trust and monitoring performance. Lee and Moray (1992) found that immediately following a failure, operator trust tended to wane. Parasuraman et al. (1993) demonstrated the potential impact of this loss of trust on monitoring performance. Collectively, these findings demonstrate that an operator's dynamic perception of system reliability impacts subsequent monitoring performance. As such, the pattern of failures for the difficult monitoring task was manipulated in the present investigation so that half of the participants experienced critical deviations limited to one system while the other half experienced an even distribution between two systems. Lee and Moray found that operator trust was reduced immediately following a failure. Consistent with their findings, it was expected that trust in systems that fail more frequently and consistently would be reduced and monitoring performance in those systems would consequently improve. Therefore, it was expected that an equal

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

distribution of failures among two systems would yield poorer monitoring performance for detecting subsequent failures compared to a pattern of failures limited to only one system.

The Effects of Vigilance

Because the present study required individuals to operate a system and monitor dynamic displays for critical events over an extended period of time, it was important to acknowledge the potential impact that a loss of vigilance would have on performance. In traditional vigilance research, operators are required to detect infrequent and unpredictable signals over long intervals. The need for research on vigilance became apparent during World War II when radar operators were consistently unable to detect targets in the water (Mackworth, 1948). Over the years, research on vigilance has generated two basic conclusions: The baseline level of operator vigilance is often lower than desired, and operator vigilance levels often decline precipitously within the first half hour of the watch (Davies & Parasuraman, 1982; Mackworth, 1948). However, most research on vigilance has been conducted using very simple tasks. Because the present study required concurrent monitoring of several different types of critical events occurring in separate display locations, it was important to consider the research on vigilance performance in complex monitoring environments.

Much of the early work on vigilance using complex displays found little or no decrement over time (Adams, Stenson, & Humes, 1961; Jerison & Wing, 1957). Researchers argued that more complex displays were sufficiently engaging to eliminate the changes in arousal that led to degraded performance over time when simpler tasks were used. However, unlike the findings of Adams et al. and Jerison and Wing, other

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

researchers did find evidence that a vigilance decrement might exist under more complex monitoring tasks. Specifically, Sanders and Ferrari (1960) and Wiener (1964) both found evidence for a vigilance decrement in tasks that required monitoring multiple displays. Parasuraman (1986) argues that the failure of early research to reveal the presence of a vigilance decrement as observed by Sanders and Ferrari and Wiener may have resulted from large individual differences in the ability to monitor complex displays and levels of performance at the outset that were already impoverished. In addition, a series of studies conducted by Howell, Johnston, and Goldstein indicated that even in the absence of a decrement in critical signal detections, a significant increase in response latencies was obtained (Howell, Johnston, & Goldstein, 1966; Johnston, Howell, & Goldstein, 1966; Johnston, Howell, & Williges, 1969). Similar results have also been found by Thackray, Bailey, and Touchstone (1979) in a simulated air traffic control task that required monitoring of several displays for changes in alphanumeric signals.

More recently, research by Grubb et al. (1995) and Molloy and Parasuraman (1996) has demonstrated evidence for a vigilance decrement for monitoring performance in complex flight simulation tasks. Specifically, Molloy and Parasuraman found that in a complex flight simulation task, operators detected a signal more frequently in the first block of the experiment than in the final block. Grubb et al. also found a vigilance decrement for operators in a complex flight simulation task. Specifically, they found that operators who had to detect deviations in multiple displays performed more poorly over time, with poorer overall detection performance for operators monitoring the greatest number of systems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Taken together, this line of research provides evidence for degraded monitoring performance with respect to both detection times and absolute detection rates in complex monitoring environments. Although performance differences observed over time in many of the earlier studies were limited to increases in response latencies (Howell, Johnston, & Goldstein, 1966; Johnston, Howell, & Goldstein, 1966; Johnston, Howell, & Williges, 1969), more recent studies by Molloy and Parasuraman (1996) and Grubb et al. (1995) have shown that individuals operating in complex task environments have greater difficulty detecting critical signals over time, supporting the presence of a vigilance decrement. Therefore, it is important to investigate further the impact of vigilance for monitoring systems that use multiple critical signals that vary in cognitive complexity, over an extended timeframe.

Accordingly, one of the primary purposes of the present investigation was to examine the impact of different levels of task complexity and multiple types of critical signals on monitoring performance. By including multiple concurrent monitoring responsibilities and manipulating the complexity of the monitoring tasks, the present study addressed the failure of previous research in providing an adequately demanding monitoring situation. Consistent with Grubb et al. (1995), it was anticipated that individuals would have greater difficulty detecting failures for a more complex monitoring task that placed higher demands on cognitive resources. In addition, those individuals operating under high reliability were expected to have greater deficiencies in performance when monitoring for the more complex type of failure. This expectation was consistent with Langer's (1989) notion of premature cognitive commitment, which suggests that an operator's initial experience with a system is reinforced over time, in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

conjunction with the already demanding nature of the more difficult monitoring task. Specifically, operators under high reliability, not expecting to experience frequent failures, were expected to have even greater difficulty detecting failures that required a greater expenditure of attentional and cognitive resources. Further, monitoring performance for the more cognitively demanding task would degrade across trials.

With respect to the pattern of failures, it was anticipated that operators who experienced an equal distribution of failures would have greater difficulty detecting subsequent failures in that system. This manipulation addressed the findings of Lee and Moray (1992) and Parasuraman et al. (1993) suggesting that the pattern of system failures could impact subsequent monitoring performance. Finally, consistent with research by Grubb et al. (1995) and Molloy and Parasuraman (1996), operators under both high and low reliability were expected to have better detection rates for all three monitoring tasks at the beginning rather than the end of each experimental session.

Trust Between Humans and Machines

Despite the obvious relationship that factors like system reliability, complacency potential, and task complexity have with respect to monitoring performance, operator trust may act as a critical moderator of monitoring performance ultimately giving rise to automation-induced complacency. To date, however, the role of trust in monitoring automated systems or automation-induced complacency has not received much empirical attention. Operator trust and/or "overtrust" is often cited as inducing complacency, but is generally treated as an anecdotal factor with little empirical support delineating its specific impact on monitoring performance (Parasuraman et al., 1993; Parasuraman & Riley, 1997). One of the primary purposes of the present study was to examine the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

impact of operator trust in automation on monitoring performance. Specifically, how would trust affect an individual's ability to monitor for failures or unanticipated states especially with increasing system experience? In addition, would operator trust interact with other complacency-related factors further degrading operator monitoring performance?

As automated systems have become both more prevalent and complex, the role of the operator has evolved from one of direct manual control to that of a supervisory controller (Wiener & Curry, 1980). As a result, many researchers have hypothesized that the concept of trust is critical for examining the interaction between humans and automation. (Muir, 1987,1989,1994; Muir & Moray, 1996). Trust in automation and other advanced decision-making aids can have two important implications. First, no matter how effective or "intelligent" the automation, if it is not trusted, it may be rejected and any potential benefits may be lost (Muir, 1987; Parasuraman & Riley, 1997). Second, automation may elicit levels of trust that are unwarranted, leading to complacency, and resulting in degraded monitoring performance (Muir, 1987; Parasuraman et al., 1993; Parasuraman & Riley, 1997). In the following sections, the foundations of trust as a construct will be discussed as well as its dynamic nature and those factors that both foster and ultimately undermine it.

Definition and Dimensions of Trust

Over the years, there have been a myriad of definitions for trust. Rotter (1980) describes trust as a generalized expectancy held by an individual that the word, promise, or written statement of other individuals or a group can be relied on. Trust has also been described as an expectation related to the subjective probability an individual assigns to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the occurrence of some set of future events (Rempel et al., 1985). Further, Rempel and Holmes (1986) regard trust as the degree of confidence an individual experiences when he or she thinks about a relationship. Although each of these definitions captures some of the singular aspects of trust, the taxonomy proposed by Barber (1983) describes trust along three dimensions, suggesting a multifaceted character. Barber's three dimensions include *persistence of natural and social laws, technically competent role performance,* and *fiduciary obligations and responsibilities*. Persistence refers to the expectation of both natural (e.g., physical and biological dimensions) and moral-social order (e.g., humankind will be good and decent). Technical competence and role performance refers to the ability of those with whom we interact in relationships to perform their roles safely and effectively. The final dimension, fiduciary obligations and responsibilities, posits that our partners in interaction will place other individual's interests before their own.

Although Barber's (1983) taxonomy was originally discussed in the context of human interaction, Muir (1987) has argued for the application of Barber's taxonomy to human-machine relationships. Specifically, she argues that our expectation of the persistence of natural laws allows humans to create mental models describing system operation and to further implement those models as the rule bases and algorithms underlying the functioning of automated systems. She argues further that the expectation of technically competent role performance is fundamental for trust between humans and machines and points to Barber's classification of technical competence in three categories: *expert knowledge, technical facility*, and *everyday routine performance*. These dimensions correspond closely with Rasmussen's (1983) taxonomy of skill, rule, and knowledge-based behavior. It is important to note with regard to technical competence,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that at any given time, a human or a machine may exhibit proficiency in only a subset of these competencies. For example, it can be expected that the average homeowner can detect a leaky faucet but that he or she might be unable to diagnose and repair the specific problem. Similarly, an automated device such as the Engine Instrumentation and Crew Alerting System (EICAS) can be expected to routinely gather data regarding engine parameters, but cannot correct problems when they are detected.

Muir (1987) suggests that Barber's (1983) third dimension of trust, fiduciary responsibility, describes situations where an operator's technical competence is exceeded by an automated system or when an automated system's operations are not well understood. Automated devices are often used because they possess greater expertise or ability in a desired domain. As such, an operator may not possess the expertise or ability to directly assess the competence of the machine, (e.g., whether the Flight Management System is correctly using GPS data for automated navigation). An operator must, therefore, rely on his or her evaluation of the system's responsibility designated as the appropriateness and effectiveness of the system's design-based intentions (Muir, 1987).

The Dynamic Nature of Trust

Within Barber's (1983) taxonomy, trust expectations are characterized as having relatively static properties. However, others argue that there are also dynamic expectations that undergo predictable changes as a result of experience in a relationship (Rempel et al., 1985). Consequently, Rempel et al.'s model has been extended to describe how human trust in automation can change over time resulting from continued system experience (Muir, 1987, 1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Rempel et al. (1985) have suggested that in the early stages of a relationship, individuals base their trust primarily on the *predictability* of another person's behaviors. Similarly, in the early stages of a human-machine relationship, an individual also judges the predictability of a machine by evaluating the consistency of its behaviors over time (Muir, 1987, 1994). The elevation of trust therefore depends upon the human's continuing ability to estimate the predictability of the machine. If at any time in the initial stages of the relationship it becomes difficult to continue making attributions about the machine's predictability, levels of trust will become diminished. Muir argues further that as trust develops, system monitoring will become reduced and consequently system knowledge will be degraded. This degraded knowledge of system functioning accompanied by increasing levels of trust is at the heart of automation-induced complacency.

As a relationship progresses, trust in another person or machine depends more upon the attribution of a *dependable disposition* (Rempel et al., 1985). This attribution can be characterized as a judgment based upon a summary of behavioral evidence that expresses the degree to which a person or machine can be relied. According to Muir (1987), the attribution of dependability is based upon perceived predictability, but with an emphasis placed on events involving risk. Therefore, to establish the dependability of a human or machine at this stage, the referent must exist in some environment, which demonstrates inherent risk, that is, where the opportunity exists to be undependable. By successfully dealing with risky situations, that individual or system generates the behavioral evidence necessary for establishing the attribute of dependability.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The final stage of development of trust between humans or between humans and machines requires the establishment of *faith* (Rempel et al., 1985). Because human behavior is not deterministic, we cannot know that an individual will remain dependable over time. The same is true of machines. Because we may base our attribution of dependability on a relatively small sample of behaviors, this sample may not be representative of future behaviors. Therefore, the uncertainty of future events requires a leap of faith on the part of the operator to come to the conclusion that a system will remain dependable. In the case of human interpersonal relationships, a referent's history of both predictability and dependability plays a large part in the development of faith. However, special weighting is given to events that demonstrate the referent's intrinsic motivations to remain in the relationship (Rempel et al., 1985). Although referent motivation has little relevance to human-machine relationships, the development of faith remains a necessary and important step in human-machine interaction. For example, given the complexity of automation and the interaction that occurs between automated subsystems in many complex environments, most processes defy a comprehensive understanding by their operators.

According to Muir (1987), because operators use these systems despite being unable to comprehend their full complexity, implies that individuals have taken some leap of faith. Faith, therefore, represents the necessary assurance that the system will remain dependable in the face of future uncertainty given the operator's incomplete or even incorrect knowledge of system functioning. Muir further stipulates that given no real analogue to human motivation with current machine/automated technology, the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

development of faith may be based mostly on predictability and dependability but may also depend upon extensive system experience.

Empirical Research on Trust between Humans and Machines

In response to the relative dearth of research surrounding the impact of trust on human-machine interaction, a number of researchers in the early 1990s (Lee & Moray, 1992; Muir, 1989, 1994; Muir & Moray, 1996) began to examine this relationship. In particular, they studied those factors that contributed to losses in trust, the process and timeframe of trust recovery, performance implications resulting from losses of trust, the impact of early and late system failures on operator trust, and whether losses of trust would generalize to trust attributions in nonrelated systems.

Conducting the first in a series of investigations, Muir (1989) focused on how operator trust would impact operator allocation of functions between manual and automated control using a simulated pasteurization control plant. In her first experiment, she found that participants were able to generate meaningful and sensitive ratings of trust in machines. Her first experiment failed, however, to demonstrate a relationship between overall trust and the total time that participants used automated control. In her second experiment, Muir was able to demonstrate a strong positive correlation between trust in an automated device and the total time it was used. The second study also yielded a strong negative relationship between overall trust in an automated device and the time spent monitoring the system that it controls. Two plausible reasons have been put forth to account for the differences between the two studies including increased specificity for the trust ratings as well as an alteration in the task and reward structure used in the second experiment (Lee & Moray, 1992). Collectively, Muir's findings demonstrate the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

viability of measuring operator trust in human-machine interaction and also represent some of the earliest empirical evidence of the effects of trust on automation-induced complacency.

Lee and Moray (1992) conducted a subsequent series of studies that examined trust between humans and machines. They also used a simulated pasteurization plant and focused on the strategies used for switching between manual and automatic control for maintaining optimal performance. Their results indicated that system performance, (i.e., the reliability of the system), was one of the primary factors impacting the development of operator trust. Parasuraman and Riley (1997) have argued that Lee and Moray's findings demonstrate how highly reliable systems can elicit operator overreliance, resulting in degraded monitoring performance. As such, operators of highly reliable systems may experience automation-induced complacency, limiting their ability to detect infrequent or unanticipated system states. In addition to demonstrating the importance of system reliability, Lee and Moray also showed that trust exhibited dynamic properties and that trust was lost and recovered over time in response to both the overall quality of system performance and in response to system failures.

In a final set of studies, Muir and Moray (1996) further validated the integrated model of trust in machines developed by Muir (1987, 1994) supporting the use of subjective ratings. Operators were able to provide ratings of trust that were sensitive to the specific properties of the automation. Muir and Moray also found that the construct of *competence* (Barber, 1983; Muir, 1987, 1994) best captured what operators have in mind when they express trust in an automated system. Because competence refers to the extent to which an automated system can perform its function properly, (i.e., system

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reliability), it is not surprising that operators use this dimension as their primary consideration when determining how much to trust a system. Finally, Muir and Moray found a strong positive correlation between an operator's level of trust and the amount of time spent in an automated mode. The authors argued that this finding bolsters the assertion that operator trust can provide a meaningful insight into the strategies that operators employ for using complex systems.

This series of studies has a number of key implications for human-automation interaction and the potential for complacency in a variety of contexts. First, the research validated the use of measures of operator trust as a predictor for trust related outcomes in human-automation interaction (Muir, 1989; Muir & Moray, 1996). Second, the authors demonstrated the existence of a negative relationship between an operator's level of trust and their monitoring performance (Muir, 1989). This finding helps to establish that higher degrees of trust in automated systems may be related to degraded monitoring performance. In addition, the results of Lee and Moray (1992) and Muir and Moray establish the impact that system reliability has on monitoring performance with high reliability systems engendering levels of reliance that may preclude effective operator monitoring.

One of the primary goals of the present research was to examine how trust in automation can impact monitoring performance. More specifically, how do the dynamic properties of trust that evolve with continuing system experience impact monitoring performance? As stated by both Rempel et al. (1985) and Muir (1987, 1994), when individuals interact with highly reliable people or systems, their levels of trust will increase over time. According to Muir, the increasing trust associated with extensive use

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of highly reliable systems may result in degraded monitoring performance. It was, therefore, necessary to examine the dynamics of trust for individuals operating highly reliable systems. If there were increases in operator trust associated with the use of highly reliable systems over time, then the short experimental timeframes used by Thackray and Touchstone (1989), Parasuraman et al. (1993), and Molloy and Parasuraman (1996) would be insufficient for demonstrating automation-induced complacency.

Therefore, in contrast with previous research on automation-induced complacency, the present study examined monitoring behavior as related to an operator's trust in automated systems. Further, this investigation involved several experimental sessions to determine the dynamic effects that trust has on monitoring performance over time. It was expected that individuals operating under higher levels of reliability would exhibit elevated levels of trust. Those individuals operating under high reliability were also expected to demonstrate increasing levels of trust over time. In addition, the pattern of failures for the more cognitively demanding task was expected to impact operator trust in that task. Specifically, operators who experienced an even distribution of failures were expected to demonstrate elevated levels of trust in engine performance relative to those who experienced consistent failures in that system. Finally, it was anticipated that operator trust would significantly predict monitoring performance. Specifically, increases in trust would predict degraded monitoring performance and further, monitoring performance would suffer more as experience with the system continued.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Purpose of the Present Study

Given the continuing trend toward greater automation within complex systems (Parasuraman & Byrne, 2003), and the characterization by Wiener and Curry (1980) that operating complex systems is primarily a monitoring task, it is critical to understand those factors that both facilitate and undermine monitoring performance. As such, the present investigation included two studies that examined the impact of system reliability, technology-related attitudes, monitoring complexity, operator trust, and system experience on monitoring performance.

The goals of the first study were fivefold. First, the impact of high and low system reliability on monitoring performance was examined using levels of reliability approximating or suggested by previous research on automation-induced complacency (Parasuraman et al., 1993; Thackray & Touchstone, 1989) allowing for a more accurate comparison of performance in the present study with previous research on automation-induced complacency as well as a more realistic generalization to real-world systems. Interactions between system reliability, technology-related attitudes, operator trust, and the number of trials were also assessed. Second, the impact of technology-related attitudes, specifically complacency potential, on an operator's ability to monitor a system was addressed. Any moderating effects that complacency potential has relative to system reliability, monitoring complexity, operator trust, and the number of experimental trials were also examined. Third, the impact of different degrees of monitoring complexity was studied. This manipulation addressed the use of simple discrete monitoring tasks from previous research on automation-induced complacency and required operators to perform a more cognitively demanding monitoring task. The interaction between monitoring

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

complexity, system reliability, technology-related attitudes, operator trust, and experimental trials was also evaluated. Fourth, the impact of operator trust in automation on subsequent monitoring of that system was examined. Although the issue of trust has been investigated with respect to humans and machines, there has been very little empirical research on how trust influences automation-induced complacency. The fifth and final goal of the first study was to determine the impact that extensive system experience has on monitoring performance. Muir (1987, 1994) has suggested that increasing system experience elicits qualitative changes in how operators interact with and monitor automation. Given the limited experimental durations used in previous research on automation-induced complacency, the extended period of operation used in this study provides a better understanding of the dynamic influence of increasing system experience on performance. The interactions among system experience, system reliability, technology-related attitudes, monitoring complexity, and trust were also examined.

The second study was specifically designed to examine the influence of technology-related attitudes, operator trust, and system experience on monitoring performance. However, system reliability and the degree of monitoring complexity were not manipulated. Instead, the second study focused on an operator's ability to detect a single automation failure over several experimental trials. Thackray and Touchstone (1989) suggested that lengthy studies with infrequent failures were necessary to adequately examine automation-induced complacency. Although research by Molloy and Parasuraman (1996) did assess the ability of operators to detect a single critical event, in the present study participants experienced several experimental sessions, some of which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

included no automation failures. As a result, the second study provides a more ecologically valid task structure and together with the findings of the first study may represent a more accurate depiction of monitoring performance in complex systems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

METHOD: EXPERIMENT 1

Participants

The participants included 32 individuals ranging in age from 20 to 41 years (M = 25.5). Twenty-seven of the participants were graduate students from the Old Dominion University Psychology Department. The sample included a comparable distribution of women and men in each of the experimental conditions with 3 men and 13 women under high reliability and 4 men and 12 women under low reliability. In addition, three of the male participants experienced a fixed pattern of digital readout deviations while four of the male participants experienced an even distribution of deviations. All participants had normal or corrected-to-normal visual acuity.

Experimental Tasks

Participants operated a suite of tasks similar to activities performed by pilots in the cockpit including a flight task and three different forms of system monitoring. The flight task, the operator's primary responsibility, required participants to compensate for disturbances in the attitude of the aircraft in order to maintain level flight. The system monitoring task was a secondary task and consisted of three separate monitoring functions: gauge monitoring, automation mode monitoring, and monitoring a digital readout.

Flight Task.

For the flight task (see Figure 1), operators were responsible for maintaining level flight. Specifically, operators were asked to keep two horizontal white lines, representing the current attitude of the aircraft relative to the ground, parallel with the artificial horizon. Deviations in the attitude were derived by summing two out of phase sine waves

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of varying amplitude. Using a joystick, the operators compensated for these deviations to maintain level flight conditions. Performance on this task was evaluated by examining the deviation from level flight ten times per second. A composite value of root mean square error (RMSE) was then calculated.

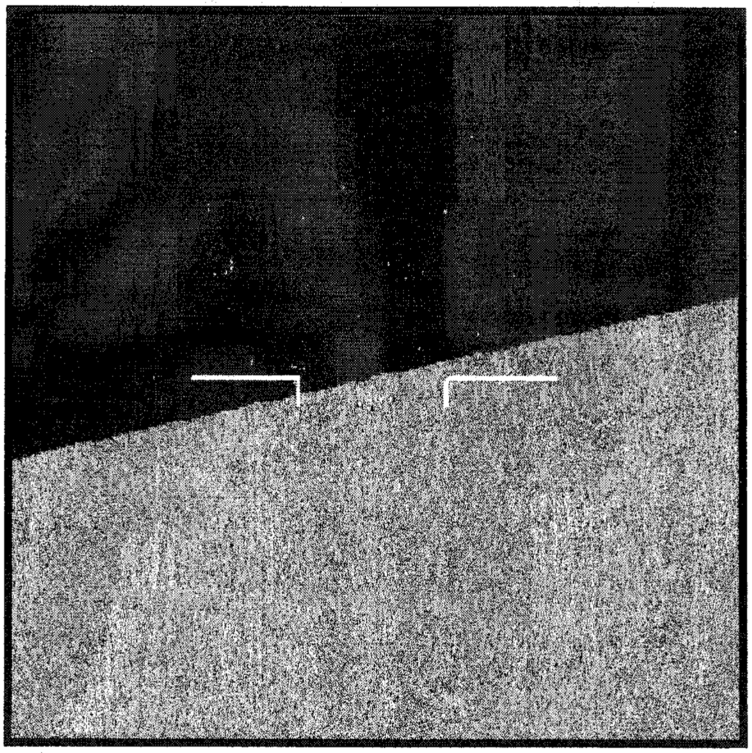

Figure 1. The primary flight task.

System Monitoring.

The monitoring task consisted of a simulated Engine Instrumentation Crew Alerting System (EICAS) display (see Figure 2). Operators were presented with three concurrent monitoring tasks. For the first, a gauge monitoring task, operators were asked to detect deviations in any of the six pointers that exceeded a critical value. Critical values were represented by two red hatch marks at each end of the circular readouts.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Under normal conditions, pointers fluctuated randomly within the normal operating range. Periodically, the gauges would move into the critical zones. Specifically, 10 critical

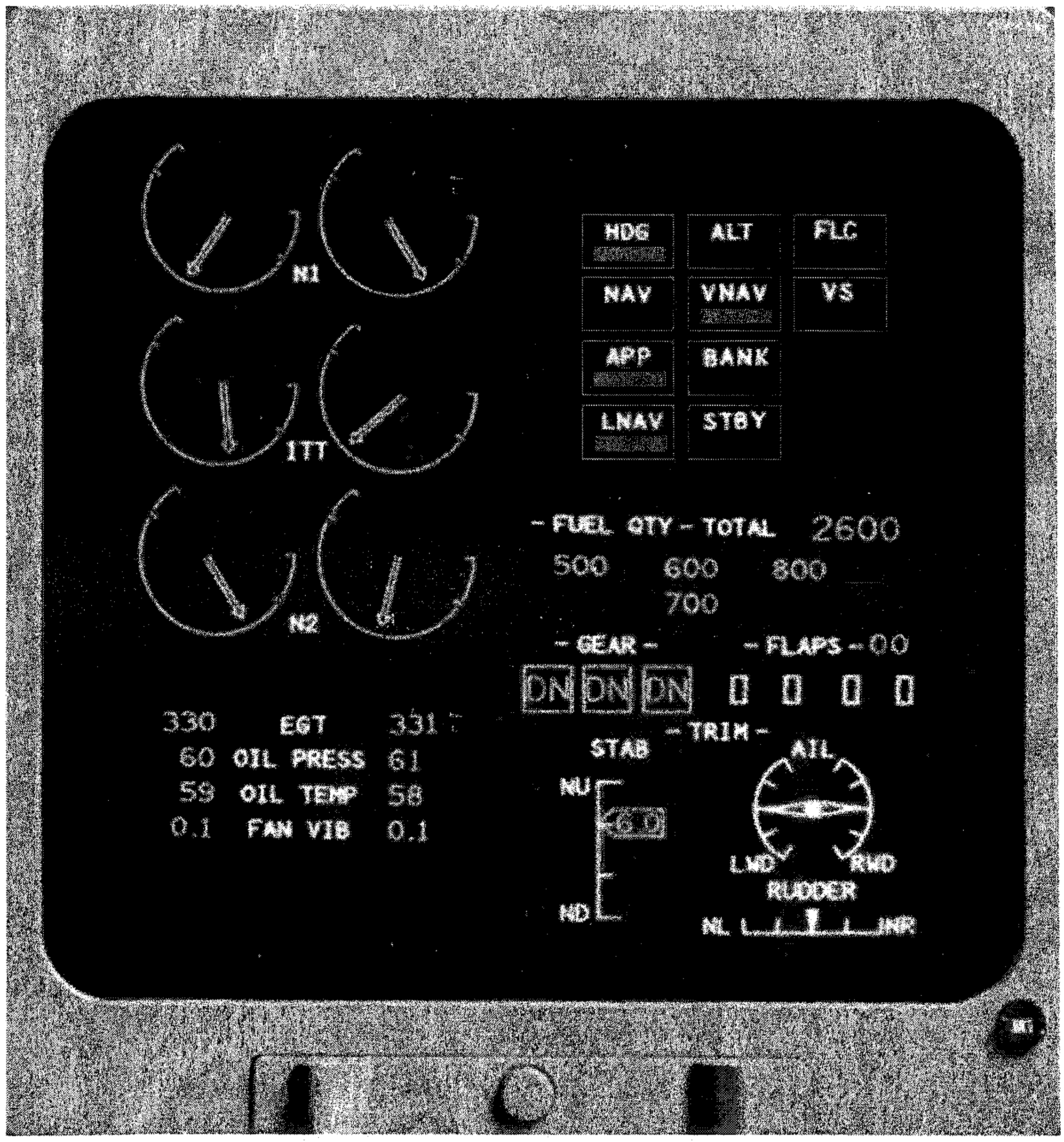

Figure 2. Simulated EICAS display with three monitoring tasks. The display includes the gauge monitoring task (upper left), automation mode monitoring task (upper right), and digital readout (bottom left).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

deviations occurred within each 10-min period of operation. Under most circumstances, these critical deviations were accompanied by an amber "Automation System 1" notification that appeared in the upper right hand corner of the EICAS display. Whenever the automation notification was illuminated, the system automatically corrected deviations in the gauge task after four s, requiring no action on the part of the operator. Occasionally, a critical deviation occurred that was not unaccompanied by the automation notification. This represented an automation failure for the gauge task. Operators were asked to respond to any deviation in the gauge task that was unaccompanied by the automation notification by pressing the "G" (in reference to the gauge task) key on the computer keyboard. Following a correct detection, the gauge returned to its normal operating range. If a participant failed to detect a critical deviation within 30 s, it was scored as a miss and the pointer reverted back to its normal range. If a participant pressed the "G" key when no critical deviation was present or when the automation notification was presented, it was counted as a false alarm.

The second monitoring task included a simulated Mode Control Panel (MCP; see Figure 2). During normal operation, four mode buttons, representing four different modes of automation, (i.e., Heading, Vertical Navigation, Approach, and Lateral Navigation), were illuminated. Similar to the gauge monitoring task, each operator experienced 10 uncommanded mode changes for each 10-min block. These mode failures were generally accompanied by an amber "Automation System 2" notification that appeared in the upper right quadrant of the EICAS display. Any change in the modes of automation accompanied by the automation notification did not require operator intervention. However, if one of the automation modes became extinguished

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the automation notification did not appear, operators were asked to correct the failure by pressing one of four keys on the computer keyboard. Specifically, operators pressed the key corresponding to the first letter of the extinguished mode, (i.e., "H" for Heading, "V" for Vertical Navigation, "A" for Approach, and "L" for Lateral Navigation). If the participant pressed the correct key within 30 s, the extinguished mode would illuminate. Otherwise, it was scored as a miss and the mode then returned to its normal status. If the participant pressed a key when there as no critical deviation or when the automation notification was present, a false alarm was recorded.

The third monitoring task required operators to monitor values on the digital readout portion of the EICAS display (see Figure 2). This task consisted of monitoring four sets of engine parameters with two values in each set, representing data from the left and right engines. Values on the left and right sides of the digital readout represented data from the left and right engines, respectively. Operators were asked to monitor four parameters including Exhaust Gas Temperature (EGT), Oil Pressure (OIL PRESS), Oil Temperature (OIL TEMP), and Fan Vibration (FAN VIB). The normal operating ranges for the first three sets of parameters were 330 ± 10 for EGT, and 60 ± 3 for the Oil Pressure and Oil Temperature. In addition, operators were told that the Fan Vibration indicator was not to exceed a value of 0.2. Operators were asked to memorize the normal operating range for each of the engine parameters. Unlike the gauge and automation mode monitoring tasks, deviations were never accompanied by any automation notifications. When a critical deviation did occur, operators were responsible for pressing a key on the computer keyboard to correct it. For each parameter, EGT, OIL PRESS, OIL TEMP, and FAN VIB, the corresponding keys were "N", "R", "S", "P",

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

respectively. These keys were chosen at random, requiring operators to memorize the appropriate input for responding to deviations in any of the four engine parameters. Following any correct detection, the engine parameter returned to its normal operating range. If the deviation went undetected for more than 30 s, it was scored as a miss. The schedule of critical deviations in the digital readout was quasi-randomly distributed throughout the four quarters of each experimental session. In addition, the pattern of critical engine deviations displayed by the digital readout was also manipulated. Half of the participants experienced critical deviations in only the left engine parameters. By contrast, the other operators experienced an even number of failures for both engines. This factor was counterbalanced across the system reliability manipulation.

For both the gauge and the automation mode monitoring tasks, the reliability of the automated system for detecting deviations was manipulated. Participants in the high reliability condition experienced a 2.0% failure rate while participants under low reliability experienced a 13.0% failure rate for each system. Specifically, for participants in the high reliability condition, the automation failed to detect 2 out of the 100 deviations. In the low reliability condition, the automated system failed to detect 13 out of 100 deviations. The level of system reliability under high reliability was chosen because the system was as reliable as possible while still allowing for a dichotomous examination of detection performance across time. By contrast, the reliability of system under low reliability was chosen to allow for a direct comparison with operator performance from research by Parasuraman et al. (1993). For the high reliability condition, failures were distributed in a quasi-random fashion among the first and fourth quarters of each experimental session. Under low reliability, the distribution of failures

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for each session occurred in a quasi-random pattern with an approximately equal distribution throughout the experiment.

Individual Difference Measures

Complacency-Potential Rating Scale

The Complacency-Potential Rating Scale (CPRS; Singh et al., 1993) was developed to measure attitudes regarding commonly encountered automated devices, (e.g., Automatic Teller Machines), that reflect the potential for automation-induced complacency (see Appendix A). A factor analysis conducted by Singh et al. for each of the scale items revealed five unique factors: Confidence-Related, Reliance-Related, Trust-Related, and Safety-Related complacency, as well as a General factor of complacency related attitudes. Singh et al. also argue that in a preliminary analysis, the instrument indicates acceptable discriminant validity based on a scale developed by Igbaria & Parasuraman (1991) examining computer use for decision-making and planning activities. In addition, the CPRS has demonstrated high levels of internal consistency ($r > .98$) as well as high levels of test-retest reliability ($r = .90$) among the items (Singh et al., 1993).

The CPRS contains 20 items, including both positive and negative statements, that utilize a 5-point Likert-type scale with response anchors ranging from *strongly disagree* (1) to *strongly agree* (5). Four of the items in the CPRS are referred to as "bogus" or "filler" items and are used as a check for response consistency. Thus, the remaining 16 test items allow for possible overall scores ranging from 16 (very low complacency potential) to 80 (very high complacency potential).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Measure of Operator Trust.

A 12-item questionnaire (see Appendix B) was developed to assess operator trust in the automated devices as well trust in overall engine performance. Each item utilized a 21-point bipolar rating scale. The instrument included four subscales, each with three items. The four subscales examined operator trust for each system with separate subscales for the left and right engines. The instrument included items such as, "Indicate how reliable you felt the automated system was at correcting any critical deviations that occurred with the gauge task", "How much do you trust the performance of the left engine based on the information from the digital readout?", and "If you were unable to monitor the automation mode portion of the display for several minutes, how confident would you be that the automation would detect any problems with the system?"

Overall ratings of trust on the operator trust questionnaire could range between 12 and 252. Operator responses from the present study ranged between 95 and 252. In addition, internal consistency for the 12-item scale as well as each of the subscales was high. Specifically, the overall reliability for the 12-item scale was $r = .94$. Internal consistency for each of the subscales were $r = .92$ for the gauge automation, $r = .89$ for the mode automation, $r = .96$ for the left engine, and $r = .96$ for the right engine.

Apparatus

Each of the experimental tasks was displayed using a Pentium IV personal computer on separate 17 in Dell E550 monitors. Participants used a standard computer keyboard along with a Microsoft Sidewinder USB joystick. The primary flight task was presented directly in front of the participant at a distance of approximately 20 in. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

monitoring task was presented to the participant's left on an adjacent display. This display was angled toward the user at 30° at a distance of 25 in.

Experimental Procedure

Each participant completed an informed consent document, after which, he or she was given the CPRS. Each participant was then provided with a set of written instructions and given a brief orientation regarding the experimental tasks during which graphical examples of each type of critical deviation were displayed on the computer. Following the orientation, participants completed a 5-min practice session that did not include any failures. After the practice session, the participants were asked if they had any questions. They then began the experimental session which lasted approximately 100 min. Upon completion, each individual completed the operator trust questionnaire

Following the first session, participants were required to return and complete two more experimental sessions, each of which was preceded by a brief reminder of the experimental instructions. Following both the second and third sessions, participants were asked to complete the same questionnaires from the first session. Following the third session, all participants were debriefed.

Experimental Design

A 2 Reliability (high or low) X 2 Pattern of Digital Readout Deviations (fixed or even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed-subjects experimental design was used with system reliability and the pattern of digital readout deviations as nested variables. Operator complacency potential and ratings of operator trust were also used as predictors. Dependent measures included detection performance, response time, the number of incorrect responses, the number of false

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alarms, operator trust, and overall RMSE on the flight task. In addition, separate analyses were performed for each level of reliability to examine monitoring performance within each session. Specifically, a 2 Block (first or second) X 3 Trial mixed and 10 Block (1-10) X 3 Trial mixed subjects experimental design was performed for high and low reliability conditions at each level of monitoring complexity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

METHOD: EXPERIMENT 2

Participants

There were nine participants in Experiment 2 including five men and four women with a mean age of $M = 22.9$ years. Five of the participants were graduate students from the Old Dominion University Psychology Department. All participants had normal or corrected-to-normal visual acuity.

Experimental Tasks

Participants operated a suite of flight tasks similar to those used in Experiment 1. The attitude correction flight task was identical. The monitoring was also similar, with one critical difference; operators experienced only a single failure across all experimental trials. Each operator received the same instructions used in Experiment 1 and was responsible for monitoring each of the three systems. However, they experienced only one failure in the automation to detect a critical deviation. This deviation occurred in the gauge monitoring task and the timing of the deviation was manipulated across trials, (i.e., occurring for each operator only in the first, second, or third trial). Therefore, of the 300 critical deviations in the gauge task over the three experimental trials, only 1 required operator intervention. This constituted a 99.7% rate of reliability. Participants experienced no automation failures in the mode monitoring task nor did they experience any critical deviations for the engine parameters on the digital readout.

Experimental Procedure

The experimental procedure for Experiment 2 was identical to the procedure used in Experiment 1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Experimental Design

Experiment 2 consisted of a 3 Automation Failure Schedule (first, second, or third trial) X 3 Trials mixed design, with the position of the single failure manipulated between individuals. In addition, data from the complacency potential questionnaire as well as the scale of operator trust were used as predictors. Dependent measures included whether the operator detected the single failure, the number of false alarms, response time for detecting the failure, the accuracy of the keyboard response, operator trust, and RMSE for the flight task.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RESULTS: EXPERIMENT 1

Monitoring Performance

Detection Performance

A 2 Reliability (high or low) X 2 Pattern of Digital Readout Deviations (fixed or even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed ANOVA was performed on the proportion of failures participants successfully detected. These effects are summarized in Table 1. Using a critical value of $\alpha = .05$, a significant

Table 1
Source of Variance for Detection Performance. R = Reliability, P = Pattern of Digital Readout Deviations, T = Trials, M = Monitoring Complexity.

Source	*Type III SS*	*df*	*MS*	*F*	*p*	η^2
R	1.610	1	1.610	3.995	0.055	0.037
P	0.009	1	0.009	0.022	N.S.	--
T	0.843	2	0.422	6.910	0.002	0.019
M	13.874	2	6.937	68.010	0.001	0.315
R X P	0.001	1	0.001	0.002	N.S.	--
R X T	0.126	2	0.063	1.033	0.365	0.003
R X M	0.401	2	0.201	1.966	0.151	0.009
P X T	0.248	2	0.124	2.033	0.142	0.006
P X M	0.017	2	0.009	0.083	N.S.	--
T X M	0.771	4	0.193	4.483	0.002	0.018
R X P X T	0.218	2	0.109	1.787	0.179	0.005
R X P X M	0.041	2	0.021	0.201	N.S.	--
R X T X M	0.300	4	0.075	1.744	0.146	0.007
P X T X M	0.105	4	0.026	0.610	N.S.	--
R X P X T X M	0.168	4	0.042	0.977	N.S.	--
S (R X P)	11.272	28	0.403			
S X T (R X P)	3.429	56	0.061			
S X M (R X P)	5.736	56	0.102			
S X T X M (R X P)	4.826	112	0.043			

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

effect for trials was found, $F(2, 56) = 6.910$. Tukey HSD posttests indicated that participants showed improved detection performance in the first trial ($M = 66.7\%$, $SD = .365$) relative to performance in the second ($M = 54.3\%$, $SD = .400$) and third ($M = 56.6\%$, $SD = .401$) trials. A significant effect was also found for monitoring complexity, $F(2, 56) = 68.010$, with detection performance differing at each level. The means for the gauge, mode, and digital readout detection rates were $M = 79.2\%$ ($SD = .385$), $M = 69.8\%$ ($SD = .345$), and $M = 28.6\%$ ($SD = .306$), respectively. In addition, a main effect for system reliability approached significance, $F(1, 28) = 3.995$, $p = .055$. The trend indicated that participants under high reliability ($M = 51.7\%$, $SD = .406$) had poorer detection performance compared to participants in the low reliability condition ($M = 66.7\%$, $SD = .363$).

A significant interaction was also found for trials and monitoring complexity, $F(4,$

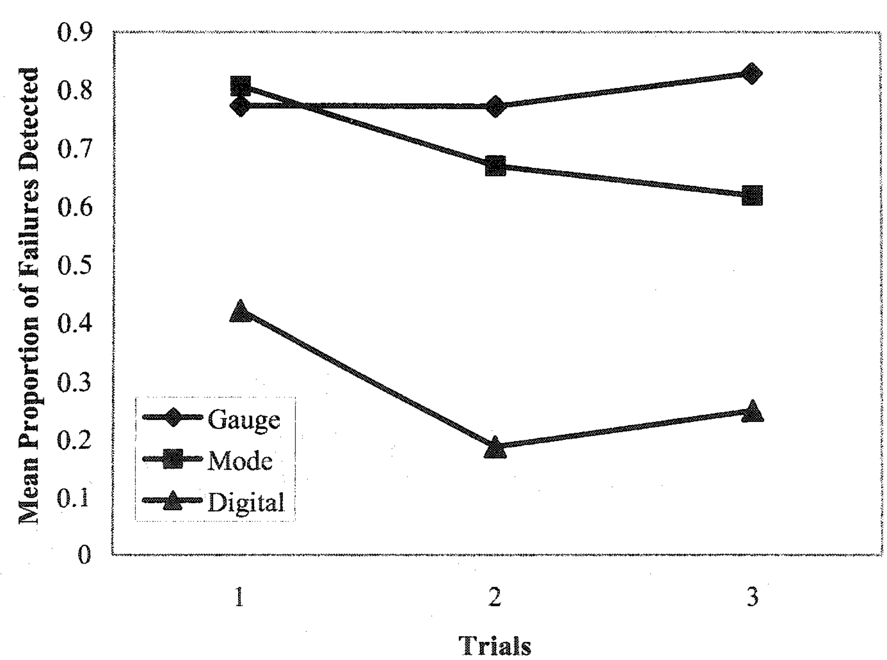

Figure 3. Detection performance as a function of trials and monitoring complexity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

112) = 4.483 (see Figure 3). In the first two trials, detection performance in the digital readout task was lower than the gauge and mode monitoring tasks which did not differ from one another. For the third trial, all levels of monitoring complexity differed from one another. In addition, posttests confirmed that across trials, detection performance for the gauge monitoring task did not differ. However, performance in the mode monitoring task declined significantly between the first and third trial. Detection performance for the digital readout task was also significantly higher in the first trial compared to the second and third trials which did not differ.

Response Time

A 2 Reliability (high or low) X 2 Pattern of Digital Readout Deviations (fixed or even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed ANOVA procedure was performed on the response times. These effects are summarized in Table 2. A significant effect for system reliability was found, $F(1, 28) = 7.077$. Those individuals under high reliability ($M = 20.431$, $SD = 9.040$) demonstrated degraded response time relative to participants in the low reliability condition ($M = 15.488$, $SD = 8.970$). The trials manipulation also yielded a main effect, $F(2, 56) = 5.848$. Posttests revealed that participants showed better response time in the first session ($M = 16.423$, $SD = 9.026$) relative to both the second ($M = 18.998$, $SD = 9.558$) and third sessions ($M = 18.457$, $SD = 9.278$). The monitoring complexity manipulation also generated a significant effect, $F(2, 56) = 81.288$. Participants monitoring the gauge task had a mean response time of $M = 13.022$ ($SD = 8.332$) while participants monitoring the mode and digital readout tasks produced mean detection times of $M = 15.657$ ($SD = 8.773$) and $M =$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

25.199 (*SD* =5.768), respectively. Posttests indicated significant differences in response times at each level of monitoring complexity.

Table 2
Source of Variance for Response Time. R = Reliability, P = Pattern of Digital Readout Deviations, T = Trials, M = Monitoring Complexity.

Source	Type III SS	*df*	*MS*	*F*	*p*	η^2
R	1759.070	1	1759.070	7.077	0.013	0.071
P	8.720	1	8.720	0.035	N.S.	--
T	353.976	2	176.988	5.848	0.005	0.014
M	7880.893	2	3940.447	81.288	0.001	0.316
R X P	8.720	1	8.720	0.035	N.S.	--
R X T	53.704	2	26.852	0.887	N.S.	--
R X M	363.113	2	181.557	3.745	0.030	0.015
P X T	79.388	2	39.694	1.312	0.278	0.003
P X M	24.884	2	12.442	0.257	N.S.	--
T X M	220.032	4	55.008	2.687	0.035	0.009
R X P X T	195.254	2	97.627	3.226	0.047	0.008
R X P X M	44.810	2	22.405	0.462	N.S.	--
R X T X M	90.271	4	22.568	1.103	0.359	0.004
P X T X M	124.966	4	31.242	1.526	0.199	0.005
R X P X T X M	79.865	4	19.966	0.975	N.S.	--
S (R X P)	6959.230	28	248.544			
S X T (R X P)	1694.713	56	30.263			
S X M (R X P)	2714.610	56	48.475			
S X T X M (R X P)	2292.583	112	20.469			

A significant interaction was found for reliability and monitoring complexity, $F(2, 56) = 3.745$ (see Figure 4). Posttests confirmed that for participants under high reliability, response times for the gauge, mode, and digital readout monitoring tasks differed from one another. For those participants under low reliability, response times for the gauge and mode monitoring tasks did not differ; however, response times for the digital readout declined compared to gauge and mode monitoring performance. In addition, participants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

under high reliability exhibited degraded response time performance for both the gauge and mode monitoring tasks relative to individuals under low reliability. No differences in

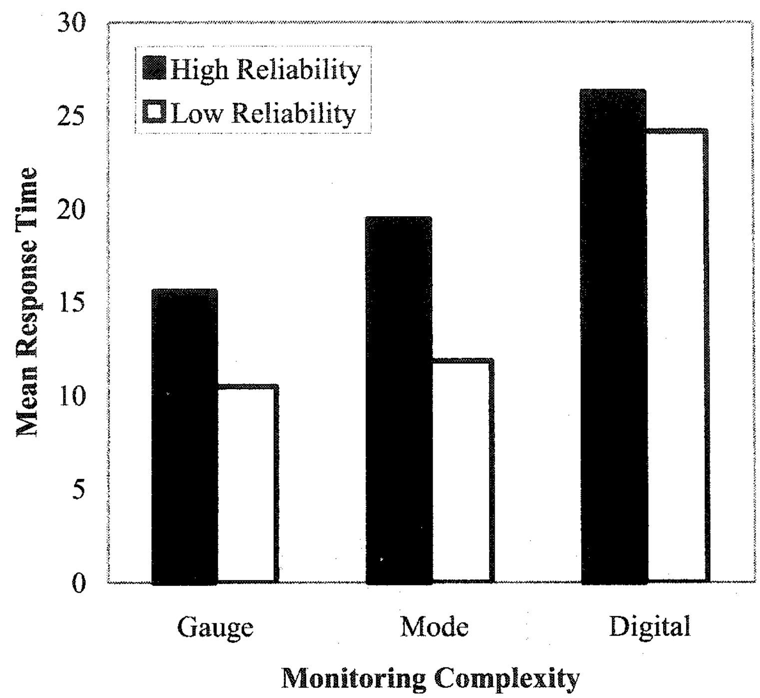

Figure 4. Response time as a function of reliability and monitoring complexity.

monitoring performance were found at either level of system reliability for the digital readout task.

A significant interaction was also found for trials and monitoring complexity, $F(2, 112) = 2.687$ (see Figure 5). In the first two trials, participants demonstrated degraded response times in the digital readout task relative to both the gauge and mode monitoring tasks which did not differ from one another. Response times in the third session were different at each level of monitoring complexity. In addition, although response times in the gauge monitoring task did not vary across trials, posttests did reveal degraded

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

response performance between the first and third trial of the mode monitoring task and the first and second trial of the digital readout task.

Finally, a significant three-way interaction was found for system reliability,

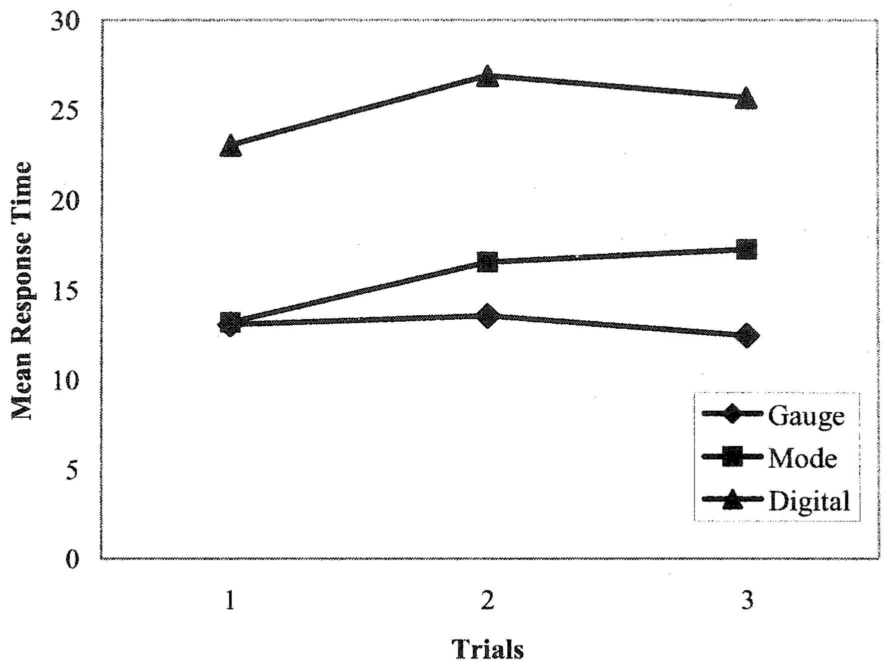

Figure 5. Response time as a function of trials and monitoring complexity.

pattern of digital readout deviations, and trials, $F(2, 56) = 3.226$ (see Figure 6). For the first trial, response times for individuals in the high-fixed condition did not differ from those in the high-even group. Likewise, response times for individuals in the low-fixed condition did not differ relative to individuals in the low-even condition. However, response times in the high-fixed condition were significantly longer than those in both low conditions. In addition, although response times in the high-even condition did not differ from those in the low-even condition, they were longer than those in the low-fixed condition. For the second trial, response times within the high and low reliability

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

conditions did not differ. Only response times for the high-fixed group were significantly longer than those of the low groups. For the third trial, response times for the high-fixed group did not differ from those of any other group. However, response times for the high-even condition were significantly longer than those of the low-fixed and low-even conditions. Finally, no differences were found across trials within any group.

Incorrect Responses and False Alarms

In addition to detection performance and response time, incorrect responses and the number of false alarms each operator committed were measured. An incorrect response was operationalized as any keyboard input, in response to an automation failure, that deviated from the appropriate responses outlined in the instructions. For example, an

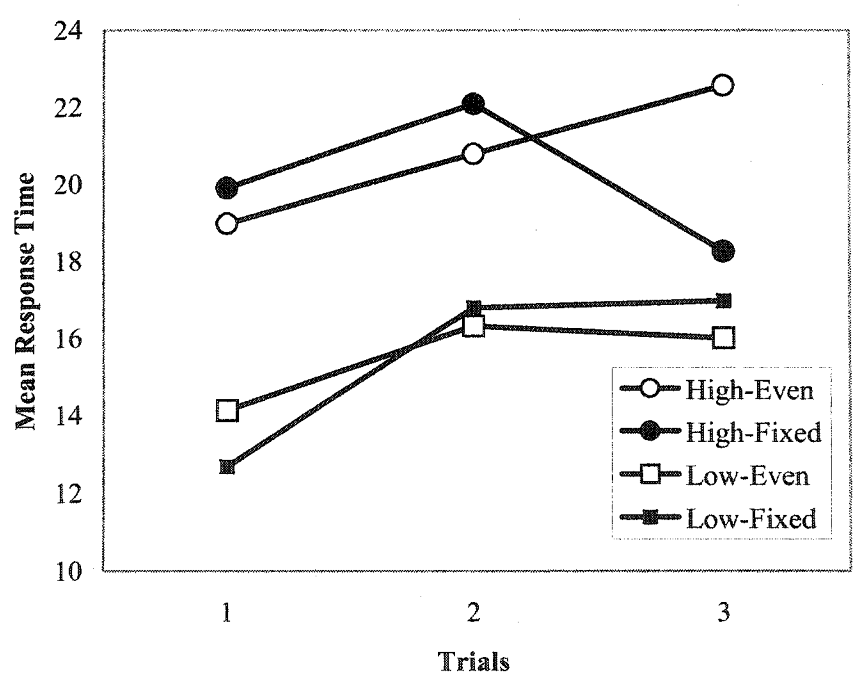

Figure 6. Response time as a function of system reliability, pattern of the digital readout deviations, and trials.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

operator might detect a failure in the gauge automation and press the spacebar. Because operators were instructed to press the "G" key in response to failures in the gauge automation, it would be deemed an incorrect response even though the operator successfully detected the failure. By contrast, a false alarm was defined as any keyboard input made by an operator when no automation failure was present.

A 2 Reliability (high or low) X 2 Pattern of Digital Readout Deviations (fixed or

Table 3
Source of Variance for False Alarms. R = Reliability, P = Pattern of Digital Readout Deviations, T = Trials, M = Monitoring Complexity.

Source	Type III SS	*df*	*MS*	*F*	*p*	η^2
R	1.389	1	1.389	0.588	N.S.	--
P	0.889	1	0.889	0.376	N.S.	--
T	34.361	2	17.181	9.602	0.001	0.071
M	13.007	2	6.504	4.173	0.021	0.027
R X P	4.014	1	4.014	1.699	0.203	0.008
R X T	0.778	2	0.389	0.217	N.S.	--
R X M	2.382	2	1.191	0.764	N.S.	--
P X T	0.194	2	0.097	0.054	N.S.	--
P X M	0.632	2	0.316	0.203	N.S.	--
T X M	18.451	4	4.613	3.584	0.009	0.038
R X P X T	2.694	2	1.347	0.753	N.S.	--
R X P X M	0.924	2	0.462	0.296	N.S.	--
R X T X M	1.701	4	0.425	0.330	N.S.	--
P X T X M	3.285	4	0.821	0.638	N.S.	--
R X P X T X M	0.868	4	0.217	0.169	N.S.	--
S (R X P)	66.139	28	2.362			
S X T (R X P)	100.194	56	1.789			
S X M (R X P)	87.278	56	1.559			
S X T X M (R X P)	144.139	112	1.287			

even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed ANOVA procedure was performed on both the number of incorrect responses and false

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alarms. No significant effects were found regarding incorrect responses. However, the analysis yielded a number of significant effects for operator false alarms. These effects are summarized in Table 3. The trials manipulation generated a significant effect for false alarms, $F(2, 56) = 9.602$. Operators committed more false alarms in the first trial ($M = .938$, $SD = 2.056$) than in either the second ($M = .250$, $SD = .580$) or the third ($M = .167$, $SD = .402$) trials. A significant effect was also found for monitoring complexity, $F(2,56) = 4.173$. Operators committed more false alarms in the gauge task ($M = .677$, $SD = 1.310$) relative to the digital readout task ($M = .167$, $SD = .451$). In addition, a significant trials and monitoring complexity interaction was observed, $F(4, 112) = 3.584$ (see Figure 7). Posttests revealed that in the first trial, participants committed more false

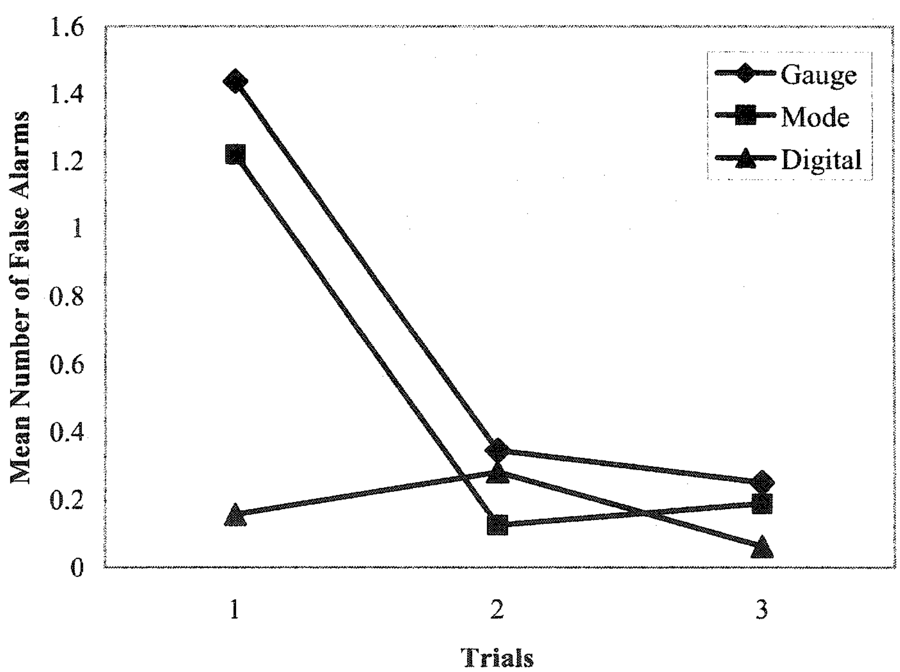

Figure 7. False alarms as a function of trials and monitoring complexity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alarms for the gauge and the mode monitoring tasks than the digital readout task. No differences were found in either the second or third sessions among any of the monitoring tasks. In addition, the number of false alarms that participants committed in the gauge and mode monitoring tasks dropped significantly between the first and the remaining trials which did not vary. No differences were observed for the digital readout task across all trials.

Intrasession Monitoring Performance

Repeated measures ANOVA analyses were used to determine whether operator performance differed within each experimental session. As noted, each session was divided into several blocks depending on the level of system reliability each operator experienced. Specifically, operators in the high reliability condition experienced two failures for each automated system. As such, a 2 Experimental Block (first or second) X 3 Trial mixed ANOVA procedure was performed on both detection performance and response time data for each level of monitoring complexity. By contrast, operators under low reliability experienced a more consistent failure rate throughout the experiment. To examine their performance within each session, a 10 Experimental Block (1-10) X 3 Trial mixed ANOVA procedure was performed on both detection performance and response time data. Results from both analyses revealed no significant effects.

Operator Trust

Group Differences for Operator Trust

Using data from the trust questionnaire, a 2 Reliability (high or low) X 2 Pattern of Digital Readout Deviations (fixed or even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed ANOVA procedure was performed. A summary

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of effects can be found in Table 4. For the following analysis, trust data from both engines were collapsed across the single level of digital readout monitoring complexity. A main effect was found for trials on operator trust, $F(2, 56) = 5.015$, with means of 48.302 ($SD = 9.432$), 51.583 ($SD = 7.639$), and $M = 50.522$ ($SD = 8.941$) for trials 1 through 3, respectively. Operator ratings of trust improved between the first and second trial, but the second and third trials did not differ. A significant main effect was also

Table 4
Source of Variance for Operator Trust. R = Reliability, P = Pattern of Digital Readout Deviations, T = Trials, M = Monitoring Complexity.

Source	Type III SS	*df*	*MS*	*F*	*p*	η^2
R	190.125	1	190.125	0.487	N.S.	--
P	175.781	1	175.781	0.450	N.S.	--
T	540.563	2	270.282	5.015	0.010	0.024
M	357.250	2	178.625	4.520	0.015	0.016
R X P	422.920	1	422.920	1.083	0.307	0.019
R X T	5.146	2	2.573	0.048	N.S.	--
R X M	281.333	2	140.667	3.559	0.035	0.013
P X T	68.396	2	34.198	0.635	N.S.	--
P X M	55.271	2	27.636	0.699	N.S.	--
T X M	266.500	4	66.625	2.319	0.061	0.012
R X P X T	12.132	2	6.066	0.113	N.S.	--
R X P X M	107.715	2	53.858	1.363	0.264	0.005
R X T X M	187.083	4	46.771	1.628	0.172	0.008
P X T X M	41.208	4	10.302	0.359	N.S.	--
R X P X T X M	39.764	4	9.941	0.346	N.S.	--
S (R X P)	10929.438	28	390.337			
S X T (R X P)	3018.042	56	53.894			
S X M (R X P)	2213.208	56	39.522			
S X T X M (R X P)	3218.000	112	28.732			

found for monitoring complexity, $F(2, 56) = 4.520$ with means of $M = 48.667$ ($SD =$ 8.532) for the gauge automation, $M = 50.417$ ($SD = 7.661$) for the mode automation, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$M = 51.354$ ($SD = 9.885$) for the engines. Posttests revealed that participants reported higher levels of trust in the performance of the engines than in the gauge automation.

A significant interaction between reliability and monitoring complexity was found for operator trust, $F(2, 56) = 3.559$ (see Figure 8). Under high reliability, operator trust did not differ for the gauge, mode, or engines. By contrast, individuals under low reliability reported lower trust in the gauge automation as compared to trust in engine performance. No differences in trust were found in the mode automation as compared to either the gauge automation or engine performance under low reliability. In addition,

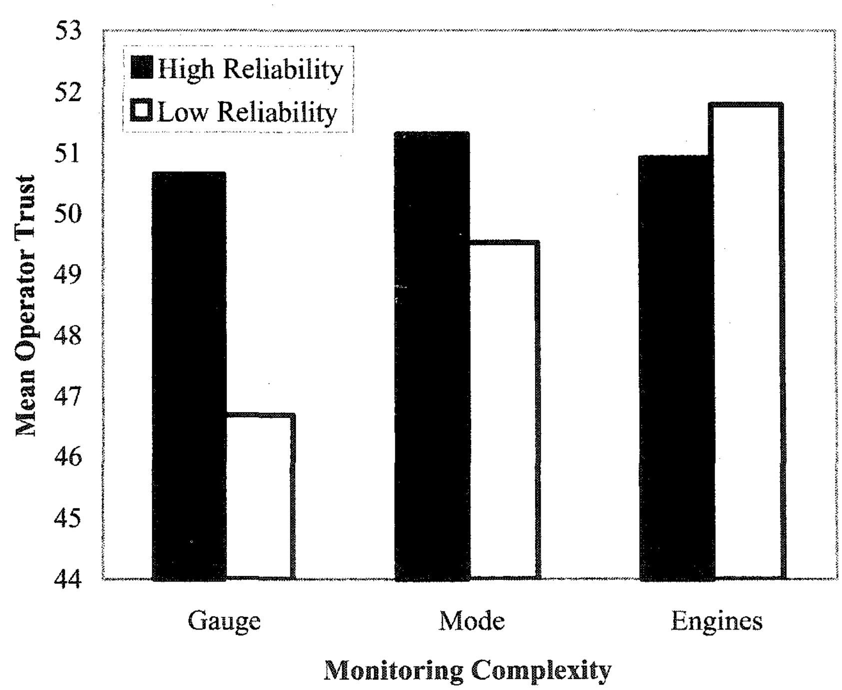

Figure 8. Operator trust as a function of reliability and monitoring complexity.

posttests confirmed that ratings of trust in the gauge automation were lower for those participants under low as compared to high reliability. Ratings of trust for the mode automation or engine performance did not differ at either level of system reliability.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Operator Trust and Monitoring Performance

A series of regression analyses were conducted to examine the impact of operator trust on monitoring performance. Trust in the gauge automation was found to significantly predict detection performance, $F(1,94) = 6.428, p = .013, R^2 = .064$, and response time, $F(1,94) = 6.493, p = .012, R^2 = .065$. Specifically, operators who placed more trust in the gauge automation exhibited degraded detection performance and response times for detecting failures in the gauge automation. Similarly, higher levels of trust in the mode automation were also found to predict degraded detection performance, $F(1, 94) = 9.544, p = .003, R^2 = .092$, and increased response time, $F(1, 94) = 8.094, p = .005, R^2 = .079$, for detecting failures in the mode automation.

With respect to operator trust in engine performance, participants provided separate ratings for the two engines. Separate ratings were necessary because participants experienced different patterns of failures in the digital readout monitoring task, (i.e., some participants experienced failures in both engines and some experienced failures only in the left engine). Regression analyses indicated that elevated ratings of operator trust in the left engine, $F(1, 94) = 22.011, p = <.001, R^2 = .190$, and right engine, $F(1, 94) = 7.866, p = .006, R^2 = .077$, predicted degraded detection performance. In addition, increased trust in the left engine, $F(1, 94) = 14.823, p = <.001, R^2 = .136$, and right engine, $F(1, 94) = 3.178, p = .078, R^2 = .033$, predicted increased response latencies. These data indicate that higher levels of trust in either of the engines led to an overall reduction in detection performance or increased response latencies for detecting deviations in the engine parameters. However, a closer examination of the data yields some interesting findings. By examining only those individuals who received an even

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pattern of failures in the digital readout task, (i.e., two failures in both the right and left engines), the predictive power of operator trust in Engine 1 and Engine 2 on both detection performance and response time is eliminated. Higher levels of trust in Engine 1 no longer predict degraded detection performance, $F(1,46) = 2.213, p = .144, R^2 = .046$, or response time, $F(1, 46) = .274, p = .603, R^2 = .006$. Similarly, elevated trust in Engine 2, also has no influence on detection performance, $F(1, 46) = 2.431, p = .126, R^2 = .050$, or response time, $F(1, 46) = .163, p = .688, R^2 = .004$, in the digital readout task. By contrast, for those individuals who experienced a fixed pattern of failures in the digital readout task, (i.e., four failures in only the left engine), the impact of operator trust, particularly in Engine 1, on both detection performance and response times was strong. Under a fixed pattern of digital readout deviations, trust in Engine 1 predicted degraded detection performance, $F(1, 46) = 23.376, p = <.001, R^2 = .337$, and response time, $F(1, 46) = 18.929, p = <.001, R^2 = .292$. Operator trust in Engine 2 for individuals under a fixed pattern of digital readout deviations also predicted degraded detection performance, $F(1, 46), = 5.537, p = .023, R^2 = .107$, but was only weakly related to increased operator response latencies, $F(1, 46) = 3.856, p = .056, R^2 = .077$.

Flight Performance

A 2 Reliability (high or low) X 2 Pattern of digital readout deviations (fixed or even) X 3 Trial X 3 Monitoring Complexity (gauge, mode, or digital readout) mixed ANOVA was performed on RMSE for flight performance. These effects can be seen in Table 5. A significant main effect was found for trials on flight performance, $F(1, 38) = 12.484$. Flight performance improved only between the first ($M = 3.299, SD = .902$) and second ($M = 2.954, SD = 1.149$) and first and third ($M = 2.759, SD = 1.150$) trials.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Complacency Potential

A series of regression analyses was conducted to determine the impact of complacency potential on both monitoring performance and ratings of operator trust. With respect to monitoring performance, higher complacency potential was associated with degraded detection performance, $F(1, 94) = 10.449$, $p = .002$, $R^2 = .100$, and increased response times, $F(1, 94) = 3.999$, $p = .048$, $R^2 = .041$, for the gauge monitoring task. Complacency potential was not predictive of monitoring performance for the mode or digital readout monitoring tasks nor did it vary as a function of system reliability or trials. Regarding operator trust, complacency potential did not predict ratings of trust in the gauge or mode automation or in the performance of the engines.

Table 5
Source of Variance for Flight Performance in Experiment 1. R = Reliability, P = Pattern of Digital Readout Deviations, T = Trials, M = Monitoring Complexity.

Source	**Type III SS**	***df***	***MS***	***F***	***p***	η^2
R	1.943	1	1.943	0.583	N.S.	--
P	0.004	1	0.004	0.001	N.S.	--
T	4.784	2	2.392	12.484	0.001	0.043
R X P	0.169	1	0.169	0.051	N.S.	--
R X T	0.348	2	0.174	0.908	N.S.	--
P X T	0.106	2	0.053	0.277	N.S.	--
R X P X T	0.445	2	0.223	1.161	0.321	0.004
S (R X P)	93.397	28	3.336			
S X T (R X P)	10.730	56	0.192			

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RESULTS: EXPERIMENT 2

Monitoring Performance

Failure Schedule, Detection Performance, and Response Time

Monitoring performance in Experiment 2 was measured by whether participants detected the single gauge automation failure and their corresponding response time. Performance was generally poor with 66.7% of participants failing to detect the single gauge automation failure.

In addition, the trial in which the single gauge failure occurred was treated as a fully counterbalanced between-subjects manipulation. Both linear and logistic regression indicated that the failure schedule had no influence on detection performance and/or response time.

A comparison of monitoring performance between Experiment 2 and Experiment 1 was also made. Because of large discrepancies in the sample sizes for the monitoring performance data in Experiment 1 and Experiment 2, homogeneity of variance tests were conducted on the detection performance and response time data. Although tests did not indicate unequal variances for the response time data, $F(2, 102) = 2.206, p = .115$, heterogeneity of variance was present in the detection performance data, $F(2, 102) = 9.260$, $p < .001$. Therefore, a more stringent level of alpha ($\alpha = .01$) was adopted for comparing monitoring data from Experiment 1 and Experiment 2. An ANOVA comparing system reliability from Experiment 1, (i.e., 87.0% for low reliability and 98.0% for high reliability) and system reliability from Experiment 2 (i.e., 99.7% reliability) revealed significant effects for both detection performance, $F(2, 102) = 9.260$, $\eta^2 = .154$, and response time, $F(2, 102) = 14.672$, $\eta^2 = .223$ (see Table 6). Participants in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Experiment 2 exhibited considerably degraded detection performance and response times compared to either level of reliability used in Experiment 1.

Table 6
Means (and Standard Deviations) for the Gauge Task at Each Level of Reliability (R1-Exp2 = 99.667%, R2-Exp1 = 98%, and R3-Exp1 = 87%) for Detection Performance, Response Time, False Alarms, Incorrect Responses, Operator Trust, and Flight Performance.

	R1-Exp2		R2- Exp1		R3-Exp1	
Detection Performance	33.333		0.729	(0.385)	0.854	(0.235)
Response Time	25.496	(8.786)	15.657	(8.996)	10.486	(6.801)
False Alarms	0.630	(2.467)	0.667	(1.492)	0.688	(1.114)
Incorrect Responses	0.000		0.000		0.292	(1.202)
Operator Trust	58.629	(3.309)	50.646	(7.413)	46.688	(9.175)
Flight Performance	3.009	(1.285)	2.862	(0.722)	3.146	(1.349)

False Alarms and Incorrect Responses

A 3 Failure Schedule (first, second or third trial) X 3 Trial mixed ANOVA was performed on the false alarm data from Experiment 2. No significant effects were found. Further, no participant in Experiment 2 committed any errors responding to the single failure in the gauge automation.

Operator Trust

Group Differences for Operator Trust

For the following analyses, only data from the trust in gauge automation subscale of the operator trust in automation questionnaire were used. A 3 Failure Schedule (first,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

second or third trial) X 3 Trial mixed ANOVA was performed on operator trust in the gauge automation and generated no significant effects.

In addition, ratings of operator trust in the gauge automation from Experiment 2 were compared with ratings from the two reliability levels used in Experiment 1. A homogeneity of variance test indicated unequal variances among the three samples, $F(2, 120) = 12.433$, $p < .001$. Therefore, a more stringent level of alpha ($\alpha = .01$) was used for comparisons of operator trust between the experiments. The ANOVA for the three levels of system reliability from both experiments on the level of operator trust in the gauge automation yielded a significant main effect, $F(2, 120) = 20.225$, $\eta^2 = .252$ (see Table 6). Operator ratings of trust in the gauge automation increased as system reliability increased.

Operator Trust and Monitoring Performance

In addition to ANOVA, linear and logistic regression were used to determine the influence of operator trust in the gauge automation on monitoring performance producing no significant effects. In contrast with findings from Experiment 1, operator trust in the gauge automation did not appear to influence detection performance or response time.

Flight Performance

A 3 Failure Schedule (first, second, or third trial) X 3 Trial mixed ANOVA was performed on RMSE for the flight performance data. These effects can be seen in Table 7. A significant main effect was found for trials, $F(2, 12) = 6.932$. Flight performance improved between the first and third trial with means of $M = 3.351$ ($SD = 1.231$), $M = 2.942$ ($SD = 1.350$), $M = 2.733$ ($SD = 1.344$) for the three trials, respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Complacency Potential

Linear and logistic regressions were computed to determine the influence of complacency potential on both monitoring performance and on operator trust in the gauge automation. Complacency potential was not predictive of detection performance or response time. However, complacency potential was marginally predictive of operator trust in the gauge automation, $F(1, 7) = 5.24, p = .056, R^2 = .428$. Higher levels of complacency potential were associated with higher levels of trust in the gauge automation.

Table 7

Source of Variance for Flight Performance in Experiment 2.

Source	**Type III SS**	***df***	***MS***	***F***	***p***	η^2
Schedule	0.281	2	0.141	0.022	N.S.	--
Trials	1.777	2	0.889	6.932	0.010	0.041
Schedule X Trials	0.239	2	0.120	0.471	N.S.	--
Subjects (Schedule)	39.084	6	6.514			
Subjects X Trials (Schedule)	1.538	12	0.128			

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DISCUSSION: EXPERIMENT 1

The goal of the present study was to examine those factors that both bolster and weaken an operator's ability to monitor automated systems. More specifically, the present investigation had five primary objectives. The first objective was to assess the effects of automation reliability on operator monitoring performance and to make comparisons with data from previous research on automation-induced complacency. Second, the present study examined the impact of technology-related attitudes, represented by complacency potential on monitoring performance. Third, the influence of task complexity on monitoring performance as well as intrasession changes were also examined. Fourth, the present study evaluated the impact of system reliability and the pattern of system failures on operator trust as well as the direct influence of operator trust on monitoring performance. Finally, the last objective was to examine the impact of increasing system experience on both monitoring performance and operator trust.

Automation Reliability and Consistency

Previous research has indicated that the reliability of a system influences operator monitoring (Lee & Moray, 1992; Muir & Moray, 1996; Parasuraman et al., 1993). Muir (1987, 1994) has also suggested that increasing experience with a system, especially a highly reliable system, can further degrade an operator's ability to monitor effectively. Accordingly, one of the primary purposes of the present investigation was to examine the influence of highly reliable systems on operator monitoring performance. In addition, the impact of overall system reliability on monitoring an unrelated system and the effects of increasing system experience were assessed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Impact of Reliability on Monitoring Performance

Data from Experiment 1 indicated that system reliability influenced the efficiency of operator monitoring. As predicted, operators who monitored a highly reliable system exhibited degraded detection performance and increased response latencies for detecting automation failures compared to individuals who monitored a system with lower reliability.

The impact of system reliability on monitoring performance may be related to operator attentional resources. As noted, Kahneman (1973) suggests that operator attentional resources are limited and that workload is a direct consequence of the disparity between task demands and the limited attentional resources available to the operator. Therefore, as the number or difficulty of tasks increases, attentional resources are depleted and operators experience increased workload and/or degraded performance. By contrast, MART posits that periods of "underload" or inactivity may also degrade operator performance (Young & Stanton, 2002). Specifically, for operators performing tasks with few demands, attentional resources shrink and performance declines as if task demands were high.

Consistent with MART, the effects of system reliability on monitoring performance from Experiment 1 may be related to depleted attentional resources. Specifically, for the present study, when operators under high reliability were first exposed to the system, one could argue that their attention was divided between the monitoring and primary flight tasks. However, as they continued to operate a system that demanded few interventions, their attentional resources became depleted and monitoring performance suffered. By contrast, participants under low reliability were frequently

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

required to make corrections in the gauge and mode automation. As such, their attentional resources remained intact and their monitoring performance remained high.

In addition, Muir (1987, 1994) has suggested that increasing system experience in highly reliable systems can further degrade monitoring performance. Therefore, it was expected that operators under high reliability would exhibit declining performance as their experience with the system increased. However, data from the present study did not reveal changes in monitoring performance over sessions as a function of system reliability. Specifically, despite overall differences in monitoring performance for operators under high and low reliability and generally degraded performance across the three experimental trials, monitoring performance at each level of reliability did not vary as a function of system experience.

One reason that performance at each level of reliability may have remained constant across time relates to the level of reliability and/or the experimental duration used in the first study. One of the main purposes of the present investigation was to use a higher level of system reliability and longer experimental sessions compared to those used in previous research on automation-induced complacency (Parasuraman et al., 1993; Thackray & Touchsone, 1989). As a result, the dynamic nature of system reliability could be examined. However, despite more realistic conditions with respect to system reliability and experimental duration, the systems used in Experiment 1 may have still been inadequate for examining monitoring performance in highly reliable systems over time. In fact, the second experiment, which used a substantially higher rate of reliability than either system in Experiment 1, was specifically designed to address this potential issue. Data from Experiment 2 and the support they provide regarding the impact of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

highly reliable systems across time will be discussed in more detail in a subsequent section.

Monitoring Performance for Unrelated Systems

In addition to the impact of system reliability on overall monitoring performance, it was expected that the reliability of the gauge and mode automation would impact monitoring performance in an unrelated system. Recall that system reliability for both the gauge and mode automation was manipulated. However, the reliability of the engines, as represented by the digital readout, remained constant. Muir and Moray (1996) found that distrust in one automated component could spread to create distrust in another automated function controlled by the same component. While it is possible that the influence of system reliability may be limited to related systems, it is also conceivable that the performance of one automated system can impact monitoring performance in an entirely separate system. However, results from the present study did not support this idea. The interaction between reliability and monitoring complexity did not indicate differences in monitoring performance for the digital readout task as a function of the reliability of the gauge and mode automation. Consistent with data from Muir and Moray, operators performed equally well under high and low reliability for detecting deviations in an unrelated system.

One potential reason that system reliability failed to influence monitoring performance may be due to the floor effect in the digital readout monitoring data resulting from generally poor performance that exhibited limited variability. Specifically, operators under high and low reliability achieved mean detection rates of only 25.0% and 32.3% and mean response times of 26.264 s and 24.134 s, respectively. Because

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

performance was so poor, the digital readout task may have been insensitive to the impact of system reliability based on the overall difficulty of that portion of the monitoring task.

Another possible explanation for why data from Experiment 1 failed to reveal an effect for system reliability on monitoring performance in an unrelated system relates to the impact of system reliability on operator attentional resources as described by MART (Young & Stanton, 2002). As will be discussed in a later section, operator reports of trust in the automation used in Experiment 1 did not differ between the high and low reliability systems. Given the generally high level of trust that operators reported in conjunction with their inability to distinguish between high and low reliability, operators may have experienced similar levels of degraded resources for monitoring for failures in the unrelated system. Therefore, monitoring performance in the digital readout task would have remained constant across the two levels of system reliability.

Complacency Potential

Related to system reliability and an operator's generalized experience with automated systems, the potential for complacency may also influence monitoring performance. Individuals maintain certain beliefs about automation that influence the way they interact with automated systems. Singh et al. (1993) suggested that these attitudes represent the culmination of user experience with automated systems and ultimately increase or decrease the potential for automation-induced complacency. Accordingly, another purpose of the present study was to examine the influence of complacency potential on monitoring performance in the context of system reliability, task monitoring complexity, and increasing system experience.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Complacency Potential and Monitoring Performance

The data from Experiment 1 provided support for complacency potential as a predictor of monitoring performance. Higher levels of complacency potential did predict degraded detection performance. Those individuals who were high in complacency potential showed reduced performance for detecting failures in the automation or deviations in the engine parameters. In addition, the relationship between complacency potential and operator response times indicated a trend in the predicted direction with increased response latencies associated with individuals higher in complacency potential.

Despite the nonsignificant relationship between complacency potential and response time, the effect of complacency potential on detection performance is arguably the more critical dependent measure, (i.e., in many situations, the ability of an operator to detect a failure is more important than the length of time needed to respond). Therefore, consistent with previous research (see Bailey et al. 2003 and Prinzel et al. 2001), results from Experiment 1 provide support for the relationship between technology-related attitudes and operator monitoring performance.

The degraded performance indicated by individuals high in complacency potential may be related to Langer's (1989) concept of premature cognitive commitment. Langer argues that operators develop attitudes regarding the efficiency of automation based on their overall experience with automated systems. Specifically, given an individual's previous experience with technology and automated systems, he or she acquires certain generalized attitudes regarding overall confidence and trust in automated systems. These attitudes then guide future behaviors and usage strategies. Therefore, the data from Experiment 1 may indicate that individuals who report a higher degree of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

trust/confidence or prefer using automation may have difficulty effectively monitoring automated systems.

Regarding the impact of system reliability and complacency potential on monitoring performance, it was expected that individuals high in complacency potential monitoring highly reliable systems would exhibit poorer performance due to degraded attentional resources. However, data from Experiment 1 indicated that the effect of complacency potential on monitoring performance was not moderated by system reliability. Individuals with high and low complacency potential performed equally well regardless of system reliability.

One possible reason that complacency potential did not vary as a function of system reliability may relate to the demands of the task. Singh et al. (1993) suggest that complacency potential by itself may not be sufficient to elicit complacent behavior. Instead, complacency potential interacts with other factors including workload, fatigue, inexperience with equipment, and poor communication to elicit poor monitoring performance. Experiment 1 was not designed to elicit extremes of these performance impairing factors. Therefore, the nature and demands of the task used in Experiment 1 may have been insufficient for revealing the effects of technology-related attitudes and system reliability on monitoring performance.

With respect to monitoring complexity, operators who possessed higher complacency potential were expected to have greater difficulty detecting failures in a cognitively demanding task. Data from Experiment 1 did indicate that complacency potential impacted performance for some forms of monitoring. Specifically, higher levels of complacency potential predicted degraded detection performance and increased

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

response latencies for the less complex monitoring task. However, higher levels of complacency potential failed to predict degraded performance in either the mode or engine monitoring tasks. Therefore, despite the more cognitively demanding nature of the engine monitoring task, those participants higher in complacency potential did not exhibit lower performance.

Consistent with data from the reliability and monitoring complexity interaction discussed in the previous section, a floor effect for operator monitoring performance may have masked the impact of complacency potential and monitoring complexity on operator monitoring performance. As noted, monitoring performance for the digital readout task was poor. As such, the digital readout monitoring task may have been too difficult to provide adequate sensitivity for investigating the impact of preexisting attitudes toward automation on monitoring performance.

With respect to system experience, because of the repetitive nature of the task, it was anticipated that increasing system experience for operators already high in complacency potential would lead to degraded monitoring performance. However, results from Experiment 1 indicated that the impact of complacency potential on monitoring performance did not change across sessions. Despite generally declining monitoring performance across the three experimental trials, performance for participants with both high and low complacency potential remained relatively consistent across the three experimental sessions.

One possible reason that data from Experiment 1 did not indicate an effect for complacency potential and system experience relates to the duration of the experiment. As noted, one of the main objectives of the present study was to use a more ecologically

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

valid task for examining automation-induced complacency. Although participants in this experiment experienced an extended period of monitoring relative to previous studies, it is possible that the effects of time on monitoring performance require even greater system experience. Given the abundant experience that operators often have with automated systems in the real world, several hours may still be inadequate for examining the subtle influence of technology-related attitudes on operator monitoring performance across time.

Another possible reason that data from Experiment 1 did not reveal an effect for complacency potential and trials relates to premature cognitive commitment (Langer, 1989). As noted, premature cognitive commitment develops when an initial system experience is reinforced over time, further confirming an operator's attitudes regarding the characteristics and efficiency of that system. Therefore, if system performance stays constant, the impact of technology-related attitudes will also remain constant regardless of how much system experience operators have. With respect to the present study, because operators experienced identical system performance over the three experimental trials, premature cognitive commitment may eliminate any potential differences in monitoring performance as a function of technology-related attitudes and increasing system experience.

In addition, although task demands, monitoring complexity, and the duration of the present study may have attenuated the relationship between complacency potential and monitoring performance, a more fundamental problem may relate to the psychometric properties of the CPRS. For example, Cronbach's alpha for Experiment 1 indicated an internal consistency of $r = .728$. This constitutes an 18% increase in measurement error compared to the level of reliability originally reported by Singh et al.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(1993). Further, the underlying factor structure for responses on the CPRS in the present study differed from those reported by Singh et al. Specifically, CPRS data from Experiment 1 generated six factors, accounting for 75.9% of the variance. By contrast, Singh et al. reported five factors which accounted for 53.2% of the overall variance. The distribution of items by factors for the present study also differed from the results reported by Singh et al. For example, Singh et al. described a *confidence-related* subscale consisting of four items. By contrast, results from the present study indicated that only two of the original four items loaded together. Similarly, for the *reliance-related* and *trust-related* subscales, although each subscale originally consisted of three items, responses from the present study indicated that only two of the original items for each subscale loaded together. Finally, for the two-item *safety-related* subscale described by Singh et al. (1993), data from the present study indicated separate factor loadings for each item.

The shift in response patterns in conjunction with significantly increased measurement error may indicate qualitative differences in how respondents in Experiment 1 interpreted the questions of the CPRS versus the original sample used to validate the measure. In fact, a number of participants in the present investigation stated that they questioned the relevance of the example technologies used in the scale and that they had difficulty relating to the items. These issues call into question the validity of the CPRS in its current form and may indicate the need for revision and revalidation of the measure. The CPRS may be able to show gross differences between groups, as indicated by its ability to predict generally degraded monitoring performance, but may lack the ability to make finer discriminations. As a result, the deficient psychometric properties

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of the CPRS may have masked the influence of technology-related attitudes on monitoring performance as a function of system reliability, trials, and monitoring complexity.

Complexity of the Monitoring Task

In addition to the impact of system reliability and operator attitudes, the complexity of the monitoring task may have also influenced operator performance. Research by Grubb et al. (1995) indicated that attentional resources become diminished and performance degrades as a function of the number of displays operators are responsible for monitoring. Therefore, monitoring performance may vary as a function of task demands. One of the primary purposes of the present study was to assess the effect of task complexity on monitoring performance and to examine further any additional effects due to the pattern of system failures or operator experience with the system. Intrasession monitoring performance was also evaluated to determine the impact of vigilance on operator monitoring in complex displays.

Task Complexity and Monitoring Performance

Data from the present study indicated that the complexity of the monitoring task heavily influenced operator monitoring performance. Consistent with Grubb et al. (1995), monitoring performance was poorest for a task that demanded greater attentional resources. Correct detections for the gauge monitoring task were nearly three times that of performance for digital readout monitoring. Likewise, performance in the mode monitoring task was more than twice as high as monitoring performance in the digital readout task. Operator performance also declined significantly for monitoring for mode automation failures compared to performance for detecting failures in the gauge

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

automation. These results indicate considerable differences in monitoring performance based on the complexity of the monitoring task and that monitoring performance is better for tasks that demand fewer attentional resources.

It was also expected that the relationship between monitoring complexity and operator performance would be moderated by system reliability. As noted, higher reliability systems are associated with degraded attentional resources. Accordingly, more complex monitoring tasks, which inherently demand greater attentional resources, in conjunction with higher system reliability should generate degraded monitoring. Response time data from Experiment 1 supported this prediction. Operators under high reliability showed degraded response time performance for both the gauge and mode monitoring tasks relative to those participants under low reliability. However, system reliability did not appear to impact monitoring performance for the digital readout task, although this finding may be a result of a floor effect in that data, (i.e., the digital readout task may have been too difficult to provide adequate sensitivity for investigating the moderating effects of system reliability).

The discrepancy between high and low reliability for the mode monitoring task was greater than for the gauge monitoring task. Operators under low reliability performed equally well in both the gauge and mode monitoring tasks. By contrast, operators under high reliability showed degraded response times for mode monitoring compared to gauge monitoring. Because the main effect for monitoring complexity indicated that the mode task required greater attentional resources than the gauge task, the greater discrepancy for mode monitoring compared to gauge monitoring under high

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reliability indicates the negative impact that high reliability has on monitoring more complex tasks.

In addition to the impact of system reliability on monitoring performance, increasing system experience was also predicted to moderate the relationship between task complexity and monitoring performance. This is consistent with the suggestion by Muir (1987, 1994) that higher levels of system experience can lead to degraded monitoring. It was expected that over time, operator attentional resources would decline due to the repetitive nature of the task which in combination with the increased demands of a more complex monitoring task would lead to further reduced performance. Data from Experiment 1 confirmed this prediction. As indicated by the main effect for trials, monitoring performance declined across the three experimental sessions. A closer inspection of the data, however, revealed that monitoring performance for the more cognitively demanding monitoring tasks declined across trials but remained constant across trials for the gauge task. Specifically, performance in the mode monitoring and digital readout tasks declined between the first and third trials. Therefore, these data indicate that extensive system experience in conjunction with more cognitively demanding monitoring tasks may severely impair an operator's ability to monitor effectively.

The Pattern of Failures and Monitoring Performance

Besides the complexity of the monitoring task, the pattern of failures that operators experience may also influence how effectively they monitor for failures. Previous research by Lee and Moray (1992) indicated that operator trust varied according to the pattern of system failures. In addition, research by Parasuraman et al. (1993)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

showed that operator performance was influenced by the schedule of failures in an automated system. Because one of the primary goals of the present investigation was to use a more ecologically valid setting for examining monitoring performance, it was imperative for operators to experience a pattern of failures indicative of real-world settings. Since these systems often fail in meaningful and systematic ways, one purpose of the present study was to determine the influence of these patterns of failure on monitoring performance. Accordingly, it was expected that monitoring performance would vary as a function of the pattern of system failures. More specifically, operators who experienced a fixed pattern of failures that occurred in related systems would exhibit better monitoring than those participants who experienced failures that were evenly distributed between two systems.

Data from Experiment 1 did not support an overall effect for the pattern of failures in the digital readout task on monitoring performance. Regardless of whether operators experienced a fixed or even distribution of failures in the digital readout task, monitoring performance remained constant.

One possible reason that the pattern of failures failed to influence monitoring performance may relate to the poor overall monitoring performance in the digital readout task. As such, operators may have failed to notice that there were two distinct failure patterns. However, a significant interaction between system reliability, pattern of digital readout failures, and trials indicated that across time, monitoring performance for those individuals under high reliability who experienced an even pattern of failures continued to degrade. By contrast, monitoring performance for individuals under low reliability or those under high reliability who experienced a fixed pattern of failures converged. This

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interaction indicates that the pattern of system failures may moderate the relationship between system reliability and the amount of experience an operator has with a given system.

These data are consistent with previous research which indicates that operators react to failures and modify their strategies and monitoring behavior accordingly (Lee & Moray, 1992; Parasuraman et al., 1993). Therefore, despite the absence of a main effect for the pattern of digital readout failures, data from the present study indicate that differences in how and when systems fail potentially interact with other factors at a higher level to influence monitoring performance.

Intrasession Monitoring Performance

As opposed to focusing only on performance across sessions, the present investigation also examined fluctuations in monitoring within each session. Previous research on vigilance in complex displays has provided tenuous results. Most early research failed to find vigilance decrements in complex displays or the effects were limited to increased response latencies as opposed to degraded detection performance (Adams et al., 1961; Jerison & Wing, 1957). By contrast, more recent research has demonstrated evidence for a vigilance decrement for operators monitoring complex displays (Grubb et al., 1995; Molloy & Parasuraman, 1996). Therefore, one of the goals of the present study was to examine vigilance performance in complex displays and also to assess the impact of monitoring for several failure types. It was predicted that operators under both high and low reliability would show better performance at the beginning than the end of each session for each of the three monitoring tasks.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Data from Experiment 1 did not indicate within-session changes in monitoring performance. For those participants under high and low reliability, performance did not vary within each session, regardless of whether operators were monitoring for failures in the gauge or mode automation or deviations in the digital readout. Further, performance did not change as a function of the experimental trial. Specifically, monitoring performance remained relatively constant from the beginning to the end of each session regardless of whether it was the participant's first, second, or third trial.

There are a number of potential explanations for the consistent monitoring performance within each session. As suggested by previous research, monitoring complex displays for multiple types of failures may be sufficiently engaging to eliminate the effects of vigilance (Adams et al., 1961; Jerison & Wing, 1957). While this suggestion does not preclude declining performance between sessions, the reduction in physiological arousal often associated with losses of vigilance within sessions is eliminated when monitoring complex systems for numerous types of failures.

Another reason that performance remained constant within sessions may relate to Langer's (1989) concept of premature cognitive commitment. Specifically, the initial level of system reliability that operators experienced guided their subsequent system monitoring strategies. Because operators encountered a constant level of reliability within each session, their initial experience was reinforced and the monitoring strategies they adopted were retained. According to Parasuraman et al. (1993), operators exhibit automation-induced complacency as a function of unchanging system reliability regardless of the absolute level of reliability. Therefore, operator monitoring performance will remain relatively stable within sessions as long as the performance of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the system holds constant. Because system performance within each session remained constant across time, operators developed premature cognitive commitment regarding the nature and efficiency of the automation. As a result, monitoring strategies and subsequent monitoring performance remained constant.

These data conflict with the expectancy theory of vigilance described by Baker (1959). Expectancy theory posits that individuals monitoring for low probability events will always underestimate the true signal probability which results in an upward shift in their response criterion. Broadbent (1971) suggested that this shift begins a "vicious cycle" that leads to degraded monitoring performance over time. However, data from Experiment 1 suggest that the influence of expectancy may be mitigated for operators monitoring complex systems. Monitoring performance in the present study did not decline within each session regardless of system reliability or task complexity. Therefore, despite the assertion by Parasuraman (1986) that operator expectancy is one of the most "potent" factors influencing vigilance, its impact may be attenuated in real-world systems where operators are often responsible for monitoring multiple systems for different kinds of signals.

Operator Trust

Unlike factors such as system reliability, complacency potential, and monitoring complexity which directly impact operator attentional resources and subsequent monitoring performance, operator trust may function as a fundamental moderator of performance in all human-automation interaction. Despite the relatively strong influence of the other factors, operator trust in automation may establish the upper bound on operator monitoring performance due to the monitoring strategies and distribution of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

attentional resources that result from whether operators undertrust, accurately trust, or overtrust automated systems.

Trust in automation has often been cited as an underlying factor that guides how efficiently operators use automation and ultimately impacts how well they monitor it (Muir, 1987; Parasuraman et al., 1993; Parasuraman & Riley, 1997). Surprisingly, however, most previous research has not empirically examined the impact of operator trust on monitoring, instead focusing more on the influence of operator trust on strategies for invoking automation or on the dynamic changes in trust that occur over time as a result of changing system reliability and/or system failures (Lee & Moray, 1992; Muir, 1987, 1994; Muir & Moray, 1996). As such, one purpose of the present study was to examine monitoring performance as a function of operator trust. Specifically, Experiment 1 assessed the dynamic nature of trust as a function of system reliability, increasing system experience, and the pattern of digital readout deviations. Additionally, Experiment 1 directly examined the influence of operator trust on monitoring performance for each of the three monitoring tasks across the three trials.

Group Differences in Operator Trust

With respect to system reliability, it was predicted that operators under high reliability would exhibit elevated levels of trust. Further, as system experience increased, trust for operators under high reliability was expected to increase. Data from Experiment 1 did not support these predictions. Operators under both high and low reliability reported equivalent trust in the automated devices. In addition, ratings of trust under both high and low reliability did not vary as a function of increasing system experience. These findings conflict with some of the previous research on trust for human-automation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interaction. Lee and Moray (1992) showed that system reliability was one of the primary factors influencing the development of operator trust. By contrast, data from Experiment 1 indicate that varying degrees of system reliability fail to elicit changes in operator trust.

One reason that system reliability may not have influenced operator trust in the present study relates to premature cognitive commitment (Langer, 1989). As noted, the initial conditions that operators experience may exert a strong influence on the style and efficiency of their subsequent interactions with automation, (i.e., systems that exhibit consistent reliability reinforce operator attitudes regarding system efficiency). Therefore, consistent with the suggestion by Singh et al. (1993), operator trust in complex systems may be influenced by system consistency and not just the absolute reliability of the automation. Because the performance of the systems that the operators experienced in the present study remained consistent across time, ratings of trust as a function of system reliability and increasing system experience might have been expected to remain constant as well.

Another reason that system reliability did not affect ratings of trust may be due to an inability of the operator to distinguish between the two levels. The effects for system reliability from previous research have resulted from systems with substantially discrepant levels of reliability. For example, research by Parasuraman et al. (1993) used a high reliability condition of 87.5% and a low reliability condition of 52.5%. By contrast, reliabilities used in the present study were more similar, (i.e., 98.0% for high reliability and 87.0% for low reliability) despite considerable differences in the absolute number of failures operators experienced. Therefore, the influence of system reliability on operator trust for systems functioning at more realistic and similar levels may differ

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from what has been found in previous research. That is, higher levels of reliability may influence operator trust in a more subtle way or require a much greater degree of system experience to impact operator trust.

With regard to the impact of failure patterns in the digital readout task on operator trust, it was expected that those individuals who experienced an even distribution of failures in both the right and left engines would report higher levels of trust in the performance of the engines. By contrast, operators who encountered a fixed distribution of failures would report lower trust in engine performance. However, data from Experiment 1 failed to support this prediction. Regardless of the pattern of failures, ratings of trust in engine performance remained the same.

Consistent with the floor effect found in the monitoring performance data for the digital readout task, operators may have been unable to discern the subtle difference in failure patterns for the digital readout. Given that operators were able to detect only 28.6% of the total deviations that occurred in the digital readout, it is unlikely they were able to discriminate between four failures in one engine and two failures in each engine. Therefore, the difficulty of the digital readout monitoring task may have precluded the pattern of failures from influencing operator ratings of trust.

Operator Trust and Monitoring Performance

One of the primary purposes of the present study was to examine how operator trust directly influences monitoring performance. It was expected that higher levels of trust would lead to degraded monitoring performance for each of the three monitoring tasks. Data from Experiment 1 supported this prediction. For both the gauge and mode monitoring tasks, elevated ratings of operator trust predicted lower detection performance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and increased response latencies. Therefore, as operator trust in the gauge or mode automation increased, corresponding monitoring performance for detecting failures decreased. Similarly, higher ratings of trust in the performance of the engines also led to degraded monitoring performance. However, by examining only those individuals who experienced an even distribution of failures, operator trust no longer predicted degraded monitoring performance. By contrast, for those participants who experienced a fixed distribution of failures, the relationship between operator trust and monitoring performance was strengthened. For operators who encountered failures in only the left engine, higher levels of trust strongly predicted degraded monitoring performance for detecting deviations in the digital readout.

Taken together, these data represent some of the first empirical support for the relationship between operator trust and monitoring performance. In general, when operator trust is high, monitoring performance is low. This supports the contention by many researchers that automation-induced complacency is heavily influenced by operator trust (Parasuraman et al., 1993; Singh et al., 1993). With respect to the digital readout task, data from Experiment 1 may indicate that the pattern of failures acts as a moderator between the level of operator trust and monitoring performance. More specifically, if operators experience a meaningful pattern of failures in an automated device, the level of trust attributed to that device may have a stronger impact on subsequent monitoring performance. By contrast, a more random pattern of failures with no discernible order has only a tenuous impact on operator trust and subsequent monitoring for that system.

One possible method for addressing further the influence of the pattern of system failures on operator trust and subsequent monitoring performance would be to manipulate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the failure schedule for a monitoring task with more salient characteristics. Specifically, manipulating the pattern of failures for the gauge task rather than the digital readout task might be more appropriate because that task discriminated among individuals according to the level of system reliability, (i.e., operators were aware of changes in the properties of the gauge automation). Although data from Experiment 2 did suggest an interaction with system reliability, the pattern of system failures, and trials, the failure pattern manipulation was expected to have a stronger influence on trust and subsequent monitoring performance. Because subtle failure patterns are more akin to what operators experience in the real world, it is critical to understand how these types of failures influence operator trust and subsequent monitoring performance. Using a more salient task would help to identify how subtle but meaningful patterns of failure impact operators monitoring performance and trust acquisition in highly reliable systems.

With respect to the impact of operator trust on monitoring performance across time, it was predicted that higher levels of trust in combination with increasing system experience would further degrade monitoring performance. However, despite an overall decline in monitoring performance across the three trials, system experience failed to interact with operator trust. Higher levels of trust did indicate lower monitoring performance but the strength of that relationship did not vary as a function of time.

One potential reason that data from Experiment 1 failed to show a relationship between operator trust and system experience on monitoring performance may be because the operators experienced consistent system performance across each of the three trials. Accordingly, their attributions of trust and the resulting influence of trust on monitoring performance may have also remained relatively stable. Thus, although higher

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

levels of trust have a negative impact on monitoring performance, the influence of trust remains constant as a function of unwavering system performance. It is also possible that these findings, again, relate to premature cognitive commitment (Langer, 1989). As discussed previously, initial experience with a system may heavily influence the subsequent style and efficiency of operator interaction with automation.

Comparison with Previous Research

The final goal of Experiment 1 was to make comparisons with previous research on automation-induced complacency. In fact, the methodology of the present research can be viewed as a culmination and extension of two previous studies with respect to system reliability, monitoring complexity, and the duration of the experiment. Thackray and Touchstone (1989) were the first to make an empirical examination of automation-induced complacency by assessing monitoring performance in an air traffic control simulator for operators with and without an automated aid. In addition, operators were required to monitor for two different types of failures, one more difficult to detect than the other. Later, Parasuraman et al. (1993) looked at the performance consequences of constant and variable system reliability for both high and low reliability systems in a complex flight simulation task.

System Reliability and Previous Research

With respect to system reliability, the present study indicated that higher levels of system reliability led to degraded monitoring performance. By contrast, the study by Parasuraman et al. (1993) failed to find a main effect for system reliability. Specifically, under their constant reliability condition, high reliability (87.5%) and low reliability (52.5%) failed to influence monitoring performance. Recall that Parasuraman et al. used

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a 10-s limit for operators to detect failures. By contrast, in the present study participants were allowed 30 s to respond. The difference in criteria used between the studies may have generated disparate results. To make a more precise comparison between the two studies, the data from Experiment 1 were reanalyzed adopting the same 10-s limit used by Parasuraman et al. The effect for system reliability in the current study was still present, $F(1, 30) = 11.29$, $\eta^2 = .091$. Using a 10-s criterion, higher reliability had a negative impact on detection performance with mean detection rates of 25.7% and 47.1% for high and low reliability conditions, respectively. By contrast, Parasuraman et al. reported mean detection rates of 28.0% for high reliability and 37.0% for low reliability.

Using the same 10-s limit for comparisons, detection rates for operators in the high reliability condition from Experiment 1 and the high reliability condition from Parasuraman et al. (1993) were nearly identical. However, data from the low reliability condition in the present study showed a 22.0% improvement in monitoring performance compared to the corresponding participants under low reliability from Experiment 1. Recall that Parasuraman et al. did not find performance differences between operators under high and low reliability. Therefore, the higher levels of system reliability used in Experiment 1 generated differences in monitoring performance between high and low reliability that the lower levels of reliability used by Parasuraman et al. failed to demonstrate.

The performance discrepancies between Experiment 1 and Parasuraman et al. (1993) may indicate substantial differences between monitoring highly and moderately reliable systems. Research by Muir (1987) and Rempel et al. (1985) has suggested that operator trust develops as a function of system reliability. Given the low levels of system

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reliability typically used in previous research, operators may never develop sufficient levels of trust to demonstrate how it influences monitoring as a function of system reliability. Data from the present study suggest that using highly reliable systems has a qualitatively different impact on trust acquisition and subsequent monitoring performance than systems used in previous research. Specifically, results from the present investigation suggest that the levels of reliability used by Thackray and Touchstone (1989) and Parasuraman et al. may be inadequate for describing how system reliability impacts an operator's ability to monitor effectively.

Monitoring Complexity and Previous Research

In addition to examining the impact of higher levels of reliability on monitoring performance, most previous research on automation-induced complacency has neglected to address the different types and levels of difficulty for monitoring failures in complex systems. For example, to measure monitoring performance, Parasuraman et al. (1993) used only a simple discrete monitoring task, (i.e., whether a pointer deviated significantly above or below a given parameter). The simplicity of this kind of monitoring task may fail to capture the complex nature of monitoring real-world systems which often require operators to monitor multiple systems for different types of failures and to detect subtle and/or unanticipated patterns of failure.

Consistent with the conversion used for the system reliability comparison between Experiment 1 and Parasuraman et al. (1993), the 30-s failure duration used in the present study was reduced to 10 s to allow a direct comparison of the monitoring complexity data from Experiment 1 and Parasuraman et al. Specifically, for Experiment 1, participants detected 53.3% of the failures in the gauge automation, a task that corresponded to the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

monitoring task used by Parasuraman et al. In addition, operators detected 43.0% of the failures in the mode automation and 12.8% of the deviations in the engine parameters. By contrast, overall monitoring performance under constant reliability for Parasuraman et al. was 32.5%. Therefore, operator performance for the simple discrete task from Experiment 1 exceeded that reported by Parasuraman et al. However, performance on the more difficult digital readout monitoring task was considerably lower than the detection rate for the simple task reported by Parasuraman et al.

The most important element of this comparison is not the performance difference for the simple monitoring task observed between the two studies. Instead, the most important issue is the degraded performance that occurred among the different levels of monitoring complexity in the present study. Operator performance in Experiment 1 indicated that the specific properties of the monitoring task have a considerable impact on the ability of operators to monitor effectively. Therefore, previous research has been remiss by not including monitoring activities that require more than basic perceptual discrimination.

In addition, data from Experiment 1 revealed significant interactions for system reliability and monitoring complexity as well as trials and monitoring complexity. Taken together, these effects illustrate how monitoring performance is impacted by different degrees of task complexity as a function of higher levels of system reliability and longer experimental durations. Specifically, higher levels of reliability have a more profound and negative influence on monitoring performance for more difficult monitoring tasks. Further, monitoring performance for more cognitively demanding monitoring tasks may continue to decline as system experience increases. As such, the lack of complexity in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

monitoring tasks in combination with low system reliability and short durations used by previous researchers on automation-induced complacency fails to accurately depict the dynamic character of operator monitoring performance. Therefore, based on diminished operator trust, limited task complexity, and short experimental durations, data from previous research may fail to reflect a realistic depiction of automation-induced complacency in complex systems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DISCUSSION: EXPERIMENT 2

The goal of the second study was to assess the ability of operators to detect a single gauge automation failure across the three experimental sessions. Both Thackray and Touchstone (1989) and Parasuraman et al. (1993) suggested that extended periods of monitoring highly reliable (99.0% or higher) systems was necessary for examining the properties of automation-induced complacency. Although research by Molloy and Parasuraman (1996) did investigate an operator's ability to detect a single automation failure, they utilized a short experimental duration. Because Muir (1987, 1994) has suggested that increasing system experience in highly reliable systems can further degrade monitoring performance, it is imperative to examine monitoring in highly reliable systems over an extended period. Accordingly, Experiment 2 examined an operator's ability to detect a single failure over several hours of monitoring. Comparisons with data from Experiment 1 and previous research were made to evaluate further the impact of system reliability on operator monitoring. Differences between Experiment 1 and 2 as a function of system reliability were also assessed. Finally, the second experiment examined the impact of trust on an operator's ability to detect a single automation failure.

Monitoring Performance

Data from Experiment 2 showed a precipitous drop in operator monitoring performance for the gauge task compared to performance under both levels of reliability in the first study. Specifically, the 99.7% reliability of the gauge automation in Experiment 2 generated only 33.3% detection rate for the single gauge automation failure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

By contrast, data from Experiment 1 indicated 72.9% and 85.4% detection rates for the gauge automation failure for participants under high and low reliability, respectively.

Consistent with previous research, these data indicate that higher levels of system reliability can negatively influence operator monitoring performance. In addition, by comparing data from Experiment 2 with data from the first study, a trend emerges that suggests that the level of reliability typically found in real-world systems may severely impair an operator's ability to monitor for unanticipated and/or infrequent system states.

With respect to the impact of reliability across time, the nonsignificant interaction between system reliability and trials from Experiment 1 indicated that the impact of high and low reliability did not change over time. However, this finding may relate to the levels of system reliability and experimental duration used in that experiment. Therefore, despite the considerable increase in reliability and experimental duration operators experienced in Experiment 1 compared to previous research, those levels may have remained inadequate for examining changes in monitoring performance across time.

By contrast, comparing the data from Experiment 2 with monitoring performance from Molloy and Parasuraman (1996) clarifies the impact of extensive system experience in conjunction with high reliability. Although Molloy and Parasuraman did use a level of system reliability comparable to the one used in Experiment 2, their experiment required operators to monitor for only a short time, (i.e., one hour of total monitoring). As a result, despite an elevated level of system reliability, operator performance remained relatively high with operators detecting approximately 65.0% of all automation failures. By contrast, the reliability used in Experiment 2 in combination with nearly six hours of monitoring yielded only a 33.3% rate of detection. In addition, Molloy and Parasuraman

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

used a 10-s failure duration. If the failure duration in Experiment 2 had used the same limit, only one participant would have detected the deviation, constituting an 11.1% rate of detection!

Because operators in the experiment by Molloy and Parasuraman (1996) had similar task responsibilities and experienced a comparable level of system performance, the primary difference between the two studies was the duration that operators were required to monitor. Given the magnitude of the drop in operator performance observed in Experiment 2, the impact of increasing system experience becomes apparent. As predicted, system reliability was influenced by increasing system experience. Therefore, the nonsignificant interaction between system reliability and trials observed in Experiment 1 may be due to the levels of reliability used in that experiment despite each being considerably higher than those used in previous research. As a result, comparing operator performance from Experiment 2 with data from the first study helps to elucidate the subtle but distinct impact that the combination of high reliability and extensive system experience can have on operator monitoring in complex systems.

Operator Trust

In contrast with data from the first study, results from Experiment 2 indicated that operator ratings of trust could be attributed to overall system reliability. For Experiment 1, operator ratings of trust remained constant regardless of the level of reliability operators experienced. By contrast, operator ratings of trust in Experiment 2 were considerably higher. Specifically, operator ratings of trust increased by 13.6% over the high and 20.4% over the low reliability systems used in the previous experiment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Consistent with research by Lee and Moray (1992), data from Experiment 2 indicated that system performance is one of the main factors influencing the development of operator trust. As system reliability increased, operator ratings of trust also increased. Therefore, the nonsignificant finding for operator trust as a function of system reliability from Experiment 1 may result from the inability of those operators to discriminate between two levels of system reliability that were relatively close. By contrast, the reliability of the system used in Experiment 2 was considerably higher and operator ratings of trust reflected that increase in system performance.

With respect to the impact of trust on monitoring performance, in contrast with results from Experiment 1, data from the second experiment did not indicate that elevated trust predicted degraded monitoring performance. However, this finding may result from the ceiling effect present in operator ratings of trust in combination with the floor effect present in the monitoring performance data. As noted, operator ratings of trust in Experiment 2 were very high compared with ratings from Experiment 1. In addition, monitoring performance in the second experiment was generally very poor. Data from Experiment 1 indicated that higher levels of operator trust predicted degraded monitoring performance. Therefore, the increase in operator ratings of trust in the automation in conjunction with the corresponding decline in monitoring performance between Experiment 1 and Experiment 2 suggests that higher levels of operator trust may increasingly degrade an operator's ability to monitor complex systems.

One possible way to show a direct connection between higher levels of operator trust and degraded monitoring performance would be to develop an operator trust questionnaire that is more sensitive to changes in trust in very high reliability systems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Because the system reliability used in Experiment 2 was much higher than what operators experienced in the first experiment, it is possible that the questionnaire used in Experiment 1 was inadequate for describing the subtle but distinct differences for operator trust in an automated system that failed only one time. More specifically, trust that operators experience when interacting with systems that exhibit reliability approaching what operators experience in the real world may be qualitatively different than the levels of trust experienced by operators using only moderately reliable systems; demanding an alternative method of examination. Therefore, the instrument used to collect operator ratings of trust from Experiment 1 may have been inappropriate for describing the subtle but potentially important changes in operator trust and any subsequent impact on monitoring performance from Experiment 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

OVERALL DISCUSSION AND CONCLUSIONS

Automated systems and computer technology are becoming increasingly sophisticated and prevalent with applications in domains as diverse as aviation, maritime operations, process control, motor vehicle operation, and information retrieval (Lee & See, 2004). As this trend continues, the need for operators to monitor automated systems for failures or unanticipated states becomes critical. However, the inherent nature of human-supervisory control and the demands it places on users may be diametrically opposed to the strengths and weaknesses of human operators.

Reason (1990) asserted that if human factors specialists wanted to conceive an activity that was completely mismatched with the strengths and weaknesses of human cognition, they might have created something similar to what is currently demanded of nuclear and chemical plant operators. Arguably, the same can be said for pilots. As was the case with the crash of EAL 401, operator reliance and trust due to high levels of system reliability may diminish an operator's ability to monitor for infrequent and/or unanticipated states.

Recently, a report from NASA's Aviation Safety Report System (ASRS) described another example of complacency due to excessive trust in highly reliable systems. The incident involved the crew of a Boeing 767-300 flying into JFK who failed to reduce their flight level according to local airspace restrictions. As a result, the aircraft violated the maximum allowable altitude for commencing their approach. Although ATC had issued an altitude change that was entered into the FMC by the first officer, the automation never engaged. Because of other preparations for landing, the pilots failed to notice that their intended descent had not initiated. As a result, the aircraft was 2000 feet

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

higher than expected upon entering the approach to JFK. The pilot who filed the report went on to say that:

> Automation in modern airliners is great and works 99.9% of the time. However, this success rates lull us into complacency, believing that the system will always do what we have programmed it to do! I still don't know why the automation remained at FL370 when the new cruise altitude was set to FL230. The failure here, however, was that we failed to notice immediately that the system was not doing what we wanted it to do.

This pilot's experience helps illustrate what many researchers have characterized as automation-induced complacency and illustrates the deleterious impact that operating highly reliable systems has on monitoring performance.

Although pilot reports of automation-induced complacency are commonly cited as causes of incidents in the ASRS, researchers have failed to use settings that allow for an adequate description of monitoring performance in real-world systems. To address this need, the present set of studies examined pilot monitoring performance in highly reliable systems over an extended period for several different types of failures. In addition, a direction comparison of operator trust and monitoring performance was made.

Results from the present set of studies indicated that realistic levels of system reliability severely impair an operator's ability to monitor effectively. Specifically, data from Experiment 1 and 2 indicate declining operator performance as a function of increasing system reliability. Further, the comparison between data from Experiment 2 and research by Molloy and Parasuraman (1996) suggests that for systems exhibiting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

levels of reliability that approach what operators experience in the real world, increasing system experience may further degrade their ability to monitor effectively.

These findings illustrate one of the main limitations of previous research on automation-induced complacency; the use of artificially low levels of system reliability. Both Parasuraman et al. (1993) and Thackray and Touchstone (1989) have acknowledged the need for examining operator monitoring in highly reliable systems. Consistent with their recommendation, the present results suggest that monitoring performance is considerably different in highly reliable systems and that it may vary as a function of both system reliability and the amount of experience operators have with the system.

Given that the reliability of the automation from Experiment 2 begins to approach what operators experience in real-world systems, the degree to which their monitoring performance was impaired is disturbing. However, even the severely degraded performance indicated by Experiment 2 may reflect an overly optimistic view of operator monitoring in highly reliable complex systems. Although three of the nine participants in Experiment 2 did successfully detect the failure, comments from the other participants indicated that they had stopped regularly monitoring the simulated EICAS display. In fact, one operator reported that while they "occasionally glanced" at the monitoring tasks in the first and second sessions, they did not monitor the systems at all in the third session, focusing exclusively on the primary flight task. Given that most commercial aircraft can travel 3-5 miles in just 30 s, a lot can happen in a very short time. Therefore, it is critical that operators immediately detect any potential problems. However, the present results suggest that in highly reliable systems, monitoring performance may become severely degraded with operators taking up to several minutes to detect deviations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In addition to system reliability, complacency potential was also shown to impact monitoring performance. Singh et al. (1993) suggested that operator attitudes toward automation and technology may increase or decrease the potential for automation-induced complacency. Data from the present studies provide partial support for their claim. In general, those operators who reported higher levels of trust, confidence, and more frequent usage of automation and technology exhibited poorer overall monitoring performance. However, the relationship between operator attitudes toward technology and monitoring performance was not moderated by system reliability, task monitoring complexity, or increasing system experience.

The complexity of the monitoring task was also shown to be one of the most important factors influencing operator monitoring performance and automation-induced complacency. Data from Experiment 1 indicated degraded monitoring performance for more cognitively demanding monitoring tasks. In addition, monitoring performance for more cognitively demanding tasks degraded further as system experience increased.

These findings illustrate one of the primary limitations of previous research on automation-induced complacency, (i.e., examining monitoring performance for simple, discrete monitoring tasks over short durations is inadequate for studying automation-induced complacency). The complex and varied nature of the monitoring tasks used in the present studies was one of the strongest influences on operator monitoring performance and represents a critical element for examining monitoring performance in complex systems.

Although more obvious failures like those operators experienced for the gauge and mode monitoring tasks are important in research on automation-induced

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

complacency, the characteristics of the digital readout task are more indicative of what operators experience in real-world settings. Specifically, the ability of operators to detect subtle patterns that are often unaccompanied by any warnings is critical. The 1992 crash of an Airbus A320 in Strasbourg France highlights this need. Specifically, when the flightcrew started their approach they selected a 3,300 foot per minute descent rate rather than the intended 3.3° flight path angle. As a result, the aircraft crashed several miles short of the runway. In this situation, the crew made a valid input which failed to trigger any warnings, leaving only a very subtle pattern of events indicating the aircraft's unintended rate of descent. Results from Experiment 1 suggest that monitoring performance for these kinds of events is very poor. In fact, almost 20% of the participants were unable to detect any of the deviations in the digital readout task across all three sessions! This result, taken together with the data from Experiment 2 regarding the impact of highly reliable automation on monitoring performance, suggests that operator detection of complex or subtle patterns may be nearly impossible.

Finally, the present set of studies revealed the direct influence of operator trust on monitoring performance. Specifically, operator trust was bolstered as a function of increasing system reliability. Further, as operator ratings of trust went up, the ability of operators to monitor effectively went down. This finding indicates a direct relationship between operator trust and degraded monitoring. Although a number of researchers have argued that monitoring performance in complex systems varies as a function of operator trust in automation, most previous research has failed to show a direct connection. In fact, a recent review of the literature on trust in automation (see Lee & See, 2004) fails to reference any empirical studies that examine monitoring performance in complex systems

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as a function of operator trust. Therefore, these data represent some of the first empirical support for a direct connection between operator trust in automation and subsequent monitoring performance and suggest that trust as a function of system reliability fundamentally influences operator monitoring performance.

Taken together, data from the present set of studies indicate that monitoring performance in more realistic settings is qualitatively different than has been indicated by previous research on automation-induced complacency. Increased system reliability, varied monitoring complexity using multiple concurrent tasks, and extensive system experience heavily influence an operator's ability to monitor effectively and as such, should be regarded as critical elements for the study of operator monitoring in complex systems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Adams, J. A., Stenson, H. H., & Humes, J. M. (1961). Monitoring of complex visual displays II: Effects of visual load and response complexity on human vigilance. *Human Factors, 3*, 213-221.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Bailey, N. B., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2003). The effects of a brain-based adaptive automation system on situation awareness: The role of complacency-potential. In *Proceedings of the 47th Annual Human Factors and Ergonomics Society Meeting* (pp. 1048-1052), Denver: Human Factors and Ergonomics Society.

Baker, C. H. (1959). Towards a theory of vigilance. *Canadian Journal of Psychology, 13*, 35-42.

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

Billings, C. E., Lauber, J. K., Funkhouser, H., Lyman, G., & Huff, E. M. (1976). *NASA aviation safety reporting system* (Technical Report No. TM-X-3445). Moffett Field, CA: NASA Ames Research Center.

Broadbent, D.E. (1971). *Decision and stress.* New York: Academic.

Comstock, J. R. Jr., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research* (Technical Report No. 104174). Hampton, VA: NASA Langley Research Center.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Davies, D. R. & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum Associates.

Farrell, S., & Lewandowsky, S. (2000). A connectionist model of complacency and adaptive recovery under automation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 395-410.

Fischoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239-260.

Grubb, P. L., Warm, J.S., Dember, W. N., & Berch, D. B. (1995). Effects of multiple-signal discrimination on vigilance performance and perceived workload. In *Proceedings of the 39th Annual Human Factors and Ergonomics Society Meeting* (pp.1360-1364), San Diego: Human Factors and Ergonomics Society.

Howell, W. C., Johnston, W. A., & Goldstein, I. L. (1966). Complex monitoring and its relation to the classical problems of vigilance. *Organizational Behavior and Human Performance, 1*, 129-150.

Hurst, K., & Hurst, L. (1982). *Pilot error: The human factors*. New York, NY: Aronson.

Igbaria, M., Parasuraman, S. (1991). Attitudes toward microcomputers: Development and construct validation of a measure. *International Journal of Man-Machine Studies, 35*, 353-373.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Jerison, H. J., & Wing, S. (1957). *Effects of noise and fatigue on a complex visual task* (Technical Report No. 57-14). Dayton, OH: US Air Force, Wright-Patterson Air Development Center.

Johnston, W. A., Howell, W. C., & Goldstein, I. L. (1966). Human vigilance as a function of signal frequency and stimulus density. *Journal of Experimental Psychology, 72*, 736-743.

Johnston, W. A., Howell, W. C., & Williges, R. C. (1969). The components of complex monitoring. *Organizational Behavior and Human Performance, 4*, 112-124.

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

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

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

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

Loeb, M., & Binford, J. (1970). Examination of some factors influencing performance on an auditory monitoring task with one signal per session. *Journal of Experimental Psychology, 83*, 40-44.

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

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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. Unpublished Doctoral Dissertation, University of Toronto, Canada.

Muir, B. M. (1994). Trust in automation: 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: II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*, 429-460.

Parasuraman, R. (1986). Vigilance monitoring and search. In K.R. Boff and L. Kaufman (Eds.), *Handbook of perception and human performance Vol. 2* (pp. 43-1 - 43-39). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

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

Prinzel, L. J., DeVries, H., Freeman, F. G., & Mikulka, P. J. (2001). *Examination of automation-induced complacency and individual difference variates* (Technical Report No. TM-2001-211413). Hampton, VA: NASA Langley Research Center.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man & Cybernetics, 13*, 257-266.

Reason, J. R. (1990). *Human error.* Cambridge, UK: Cambridge University Press.

Rempel, J. K., & Holmes, J. G. (1986). How do I trust thee? *Psychology Today, 20*, 28-34.

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

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*, 1-7.

Sanders, A. F., & Ferrari, G. A. (1960). *A neglected factor in vigilance* (Technical Report IZF No. 4). Soesterberg, Netherlands: Institute for Perception.

Sarter, N. B., & Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*, 5-19.

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

Thackray, R. I., Bailey, J. P., & Touchstone, R. M. (1979). The effect of increased monitoring load on vigilance performance using a simulated radar display. *Ergonomics, 22*, 529-539.

Thackray, R. I., & Touchstone, R. M. (1989). Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviation, Space, and Environmental Medicine, 60*, 744-748.

Wiener, E. L. (1964). Multiple channel monitoring. *Ergonomics, 7*, 453-460.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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

Wiener, E. L. (1988). Cockpit automation. In E.L. Wiener and D.C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Diego, CA: Academic Press.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23,* 995-1011.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*, 365-375.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX A

1. Manually sorting through card catalogs is more reliable than computer-aided searches for finding items in a library.

2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.

3. People save time by using automatic teller machines (ATMs) rather than a bank teller for banking transactions.

4. I do not trust automated devices such as ATMs and computerized airline reservation systems.

5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.

6. I feel safer depositing my money at an ATM than with a human teller.

7. 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.

8. People whose jobs require them to work with automated systems are lonelier than people who not have work with such devices.

9. Automated systems used in modern aircraft, such as automatic landing systems, have made air journeys safer.

10. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX A (CONT.)

11. Automated devices used in aviation and banking have made work easier for both employees and customers.

12. I often use automated devices.

13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.

14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.

15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass police radar speed-trap in case the automatic control is not working properly.

16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.

17. 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.

18. Work has become more difficult with the increase of automation in aviation and banking.

19. I do not like to use ATMs because I feel that they are sometimes unreliable.

20. I think that automated devices used in medicine, such as CAT scans and ultrasound, provide very reliable medical diagnosis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX B

1. Indicate how reliable you felt the automated system, represented by "Automation System 1", was at correcting any critical deviations that occurred with the gauge task.

2. If you were unable to monitor the gauges portion of the display for several minutes, how confident would you be that the automated system would detect any deviations that occur?

3. How much do you trust the automation to correct deviations in the gauge task?

4. Indicate how reliable you felt the automated system, represented by "Automation System 2", was at correcting any critical deviations that occurred with the mode of automation task.

5. If you were unable to monitor the automation mode portion of the display for several minutes, how confident would you be that the automated system would detect any problems with the system?

6. How much do you trust the automation to correct deviations in the mode task?

7. Indicate how reliable you felt the left engine was based on the information from the digital readout portion of the display.

8. If you were unable to monitor the digital readout portion of the display for several minutes, how confident would you be that no critical deviations would occur with the left engine?

9. How much do you trust the performance of the left engine based on the information from the digital readout?

10. Indicate how reliable you felt the right engine was based on the information from the digital readout portion of the display.

11. If you were unable to monitor the digital readout portion of the display for several minutes, how confident would you be that no critical deviations would occur with the right engine?

12. How much do you trust the performance of the right engine based on the information from the digital readout?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VITA

NATHAN R. BAILEY

Department of Psychology
Old Dominion University
Norfolk, VA 23529
Work: 757-683-4439
Fax: 757-683-5087
Email: nbail001@odu.edu

Home
1308 Stockley Gardens
Apartment 102
Norfolk, VA 23517
Home: 757-624-1305
Mobile: 757-574-8461

EDUCATION

B.A. (Psychology), Ohio University, 1998
M.S. (Psychology), Old Dominion University, 2001

HONORS AND APPOINTMENTS

Dominion Scholar Fellowship (April, 1999). Old Dominion University, Norfolk, VA.
Outstanding Student Poster Contribution (October, 2000). The Human Performance, Situation Awareness, and Automation Conference, Savannah, GA.
Aerospace Graduate Research Fellowship (May, 2003). Virginia Space Grant Consortium, Hampton, VA.
National Aeronautics and Space Administration GSRP Fellowship (May, 2003). NASA Langley Research Center, Hampton, VA.

SELECTED PUBLICATIONS

Bailey, N.R. & Scerbo, M.W. (2001). Looking for moving or stationary targets in dynamic displays. Paper presented at the annual meeting of the Human Factors and Ergonomics Society, Minneapolis, MN.

Bailey, N.R. & Scerbo, M.W. (2002). The horizontal-vertical velocity illusion: Implications for the design of dynamic displays. Paper presented at the annual meeting of the Human Factors and Ergonomics Society, Baltimore, MD.

Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., & Scott, L.A. (2003). A brain-based adaptive automation system and situation awareness: The role of complacency potential. Paper presented at the annual meeting of the Human Factors and Ergonomics Society, Denver, CO.

Catanzaro, J.M., Scerbo, M.W., McKenzie, F.D., Phillips, M.A., Bailey, N.R., & Loftin, R.B. (2003). A virtual environment for training military checkpoint guards. Poster presented at the annual meeting of the Human Factors and Ergonomics Society, Denver, CO.

Loftin, R.B., Scerbo, M.W., McKenzie, R., Catanzaro, J.M., Bailey, N.R., Phillips, M.A., & Perry, G. (2003). Training in peacekeeping operations using virtual environments. In *Advanced Technologies for Military Training* (Technical Report RTO-MP-HFM-101-03). Neuilly-sur-Seine Cedex, France: NATO-RTO.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.