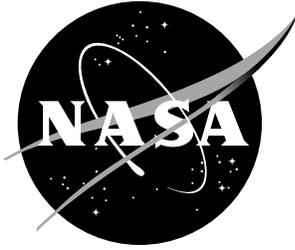


NASA/TM-2002-211925



The Relationship of Self-Efficacy and Complacency in Pilot-Automation Interaction

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

September 2002

The NASA STI Program Office ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

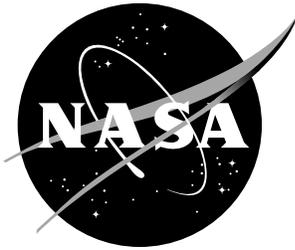
- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.
- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at <http://www.sti.nasa.gov>
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA STI Help Desk at (301) 621-0134
- Phone the NASA STI Help Desk at (301) 621-0390
- Write to:
NASA STI Help Desk
NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320

NASA/TM-2002-211925



The Relationship of Self-Efficacy and Complacency in Pilot-Automation Interaction

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

September 2002

Available from:

NASA Center for AeroSpace Information (CASI)
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161-2171
(703) 605-6000

ABSTRACT

Pilot “complacency” has been implicated as a contributing factor in numerous aviation accidents and incidents. Complacency has been defined as self-satisfaction that can result in non-vigilance based on an unjustified assumption of satisfactory system state. The term has become more prominent with the increase in automation technology in modern cockpits and, therefore, research has been focused on understanding the factors that may mitigate its effect on pilot-automation interaction. The study examined self-efficacy, or self-confidence, of supervisory monitoring and vigilance performance and the relationship of complacency and strategy of pilot use of automation for workload management. The results showed that self-efficacy is a “double-edged” sword by reducing the potential for automation-induced complacency while limiting workload management strategies and increasing other hazardous states of awareness.

TABLE OF CONTENTS

Abstract.....	iii
Table of Contents	v
Introduction	1
Automation-Induced Complacency.....	2
Trust in Automation and Complacency.....	4
Self-Efficacy and Complacency.....	4
Objectives of Present Study.....	5
Method	6
Participants	6
Baseline Task	6
Experimental Task	7
Tracking Task	7
System Monitoring	7
Fuel Management	8
Experimental Design	9
Dependent Measures	10
RMSE.....	10
A'	10
NASA-TLX	10
Experimental Procedure	10
Results	11
A'.....	11
RMSE	12
NASA Task Load Index.....	12
Discussion	13
References	15

INTRODUCTION

Automation refers to "... systems or methods in which many of the processes of production are automatically performed or controlled by autonomous machines or electronic devices (Billings, 1997, p. 7)." Billings stated that automation is a tool, or resource, that allows the user to perform some task that would be difficult or impossible to do without the help of machines. Therefore, automation can be conceptualized as a process of substituting some device or machine for a human activity. (Parsons, 1985). The dramatic increase in technology has significantly impacted all aspects of our daily lives. The Industrial Revolution ushered in an era of untold innovation that has not only made life easier and safer, but has also provided much more leisure time. One need only imagine washing one's clothes on a washing board, something considered an innovation during the early 1900's, to see how automation has transformed how we see ourselves and our place in the world. Automation has become so pervasive that many devices and machines are not even considered by most people to be "automated" anymore. Others, such as the modern airplane, however, do not escape visibility so easily. Wiener and Curry (1980), and Wiener (1989) noted that avionics has provided not only a dramatic increase in airline capacity and productivity coupled with a decrease in manual workload and fatigue, but also more precise handling, relief from certain routine operations, and more economical use of airplanes. Unlike the washing machine, the increased automation in airplanes and air navigational systems, however, has not developed without costs.

The invention of the transistor in 1947 and the subsequent miniaturization of computer components have enabled widespread implementation of automation technology to almost all aspects of flight. The period since 1970 has witnessed an explosion in aviation automation technology. The result has been a significant decrease in the number of aviation incidents and accidents. However, there has also been a corresponding increase in the number of errors caused by human-automation interaction; in other words, those caused by "pilot error." In 1989, the Air Transport Association of America (ATA) established a task force to examine the impact of automation on aviation safety. The conclusion was that,

"During the 1970s and early 1980s...the concept of automating as much as possible was considered appropriate. The expected benefits were a reduction in pilot workload and increased safety...Although many of these benefits have been realized, serious questions have arisen and incidents/accidents have occurred which question the underlying assumption that the maximum available automation is ALWAYS appropriate or that we understand how to design automated systems so that they are fully compatible with the capabilities and limitations of the humans in the system" (Billings, 1997 p. 4).

The August 16, 1987 accident at Detroit Metro airport of a Northwest Airline DC9-82 provides an example of how automation has transformed the role of pilots. The airplane crashed just after take-off en route to Phoenix. The airplane began rotation at 1,200 ft from the end of the 8,500 ft runway, when its wings rolled to the left and then to the right. The wings collided with a light pole located 0.5 miles beyond the end of the runway. One hundred and fifty-four people died in the crash with only one survivor.

For a plane to be properly configured for take-off, the flaps and slats on the wings must be fully extended. The National Transportation Safety Board (NTSB) report attributed the accident to the non-use of the taxi checklist to insure that the flap and slats of the wings were extended. The take-off warning system was cited as a contributing factor because it was not

functioning and failed to warn the crew that the plane was not ready for take-off. The airplane's stall protection system announces a stall and will perform a stick pusher maneuver to correct for the problem. However, autoslat extension and poststall recovery are disabled if slats are retracted. In addition, the tone and voice warning of the stall protection system are automatically disabled in flight by nose gear extension (Billings, 1997; NTSB, 1998). Originally, pilots manually extended the flaps and slats, performed any maneuvering needed if a stall did occur with the airplane, and were responsible for the various other tasks needed for take-off. Due to the increase in automation of the cockpit, however, they now depend on the automation to perform the pre-flight tasks reliably and without incident. Pilots have now been delegated to the passive role of monitoring the automation and are to interfere in its processes only in emergency situations.

The example above illustrates a concept known as "hazardous states of awareness" (HSA; Pope & Bogart, 1992). Pope and Bogart coined the term to refer to phenomenological experiences, such as daydreaming, "spacing out" from boredom, or "tunneling" of attention, reported in aviation safety incident reports. Hazardous states of awareness such as preoccupation, complacency, and excessive absorption in a task, and the associated task disengagement have been implicated in operator errors of omission and neglect with automated systems (Byrne & Parasuraman, 1996). The 1987 Detroit accident was caused partly by the crew's complacent reliance on the airplane's automation to configure for take-off and failure to confirm the configuration with the use of the taxi checklist (Billings, 1997).

Automation-Induced Complacency

Wiener (1981) defined *complacency* as "a psychological state characterized by a low index of suspicion." Billings, Lauber, Funkhouser, Lyman, and Huff (1976), in the Aviation Safety Reporting System (ASRS) coding manual, defined it as "self-satisfaction, which may result in non-vigilance based on an unjustified assumption of satisfactory system state." The condition is surmised to result when working in highly reliable automated environments in which the operator serves as a supervisory controller, monitoring system states for the occasional automation failure. It is exhibited as a false sense of security, which the operator develops while working with highly reliable automation; however, no machine is perfect and can fail without warning. Studies and ASRS reports have shown that automation-induced complacency can have negative performance effects on an operator's monitoring of automated systems (Parasuraman, Molloy, & Singh, 1993).

Although researchers agree that complacency continues to be a serious problem, little consensus exists as to what complacency is and the best methods for measuring it. Nevertheless, after considering the frequency with which the term "complacency" is encountered in the ASRS and analyses of aviation accidents, Wiener (1981) proposed that research begin on the construct of complacency so that effective countermeasures could be developed.

One of the first empirical studies on complacency was Thackray and Touchstone (1989) who asked participants to perform a simulated ATC task either with or without the help of an automated aid. The aid provided advisory messages to help resolve potential aircraft-to-aircraft collisions. The automation failed twice per session, once early and another time late during the 2-hr experimental session. These researchers reasoned that complacency should be evident and, therefore, participants would fail to detect the failures of the ATC task due to the highly reliable nature of the automated aid. However, although participants were slower to respond to the initial failure, reaction times were faster to the second automated failure.

Parasuraman, Molloy and Singh (1993) reasoned that participants in the Thackray and Touchstone (1989) experiment did not experience complacency because of the relatively short

experimental session and because the participants performed a single monitoring task. ASRS reports involving complacency have revealed that it is most likely to develop under conditions in which the pilot is responsible for performing many functions, not just monitoring the automation involved. Parasuraman et al. (1993) suggested that in multi-task environments, such as an airplane cockpit, characteristics of the automated systems, such as reliability and consistency, dictate how well the pilot detects and responds to automation failures. Langer (1989) developed the concept of premature cognitive commitment to help clarify the etiology of automation-induced complacency. According to Langer,

When we accept an impression or piece of information at face value, with no reason to think critically about it, perhaps because it is irrelevant, that impression settles unobtrusively into our minds until a similar signal from the outside world – such as a sight or sound – calls it up again. At that next time it may no longer be irrelevant, most of us don't reconsider what we mindlessly accepted earlier.

Premature cognitive commitment develops when a person initially encounters a stimulus, device, or event in a particular context; this attitude or perception is then reinforced when the stimulus is re-encountered in the same way. Langer (1989) identified a number of antecedent conditions that produce this attitude, including routine, repetitious, and extremes of workload; these are all conditions present in today's automated cockpit. Therefore, automation that is consistent and reliable is more likely to produce conditions in multi-task environments that are susceptible to fostering complacency, compared to automation of variable reliability.

Parasuraman, Molloy and Singh (1993) examined the effects of variations in reliability and consistency on user monitoring of automation failures. Participants were asked to perform a manual tracking, fuel management, and system-monitoring task for four 30-minute sessions. The automation reliability of the system-monitoring task was defined as the percentage of automation failures that were corrected by the automated system. Participants were randomly assigned to one of three automation reliability groups, which included: constant at a low (56.25%) or high (87.5%) level or a variable condition in which the reliability alternated between high and low every ten minutes during the experimental session. Participants exhibited significantly poorer performance using the system-monitoring task under the constant-reliability conditions than under the variable-reliability condition. There were no significant differences between the detection rates of the participants who initially monitored under high reliability versus those who initially monitored under low reliability. Furthermore, evidence of automation-induced complacency was witnessed after only 20 minutes of performing the tasks. Parasuraman et al. (1993) therefore concluded that the consistency of performance of the automation was the major influencing factor in the onset of complacency regardless of the level of automation reliability.

Singh, Molloy, and Parasuraman (1997) replicated these results in a similar experiment, which examined whether having an automated task centrally located would improve monitoring performance during a flight-simulation task. The automation reliability for the system-monitoring task was constant at 87.5% for half the participants and variable (alternating between 56.25% and 87.5%) for the other half. The low constant group was not used in this study because participants in previous studies were found to perform equally poorly in both constant reliability conditions. A constant high level of reliability was used instead because complacency is believed to most likely occur when an operator is supervising automation that he or she perceives to be highly reliable (Parasuraman et al., 1993). Singh and his colleagues found the monitoring of automation failure to be inefficient when reliability of the automation was constant but not when it was variable, and that locating the task in the center of the computer screen could not prevent these failures. These results indicate that the automation-induced complacency effect discovered

by Parasuraman et al., is a relatively robust phenomenon, which is applicable to a wide variety of automation reliability schedules. The poor performance in the constant-reliability conditions in both research studies may be a result of the participant's premature cognitive commitment or perceived trust in the automation to correct for system failures.

Trust in Automation and Complacency

Automation reliability and consistency have been shown to impart trust and confidence in automation (Lee & Moray, 1994; Muir, 1987; Muir & Moray, 1996). Muir (1994) defines trust in human-machine relationships as, "Trust (T) being a composite of three perceived expectations: the fundamental expectation of persistence (P); technically competent performance (TCP) which includes skill-, rule-, and knowledge- based behaviors, as well as reliability and validity of a referent (machine); and to fiduciary responsibility (FR) of the automation."

The specific expectation of technically competent role performance is the defining feature of trust between humans and machines. Barber (1983) identified three types of technical competence one may expect from another person or a machine: expert knowledge, technical facility, and everyday routine performance. Muir (1987) suggests that a human's trust in a machine is a dynamic expectation that undergoes predictable changes as a result of experience with the system. In early experiences a person will base his or her trust upon the predictability of the machine's recurrent behaviors. Automation reliability may instill trust and confidence in the automated system. However, trust in the automation often declines after an automation malfunction or failure, but will recover and increase as long as there are no further malfunctions. Therefore, long periods without failure also may foster poor monitoring of the automation (Lee & Moray, 1992; Riley, 1989).

Sheridan and Farrell (1974) first expressed concern about the changing roles in the modern cockpit, in which the role of a pilot changed to a supervisory controller of automation significantly transforming pilot-machine interaction. Muir (1989) confirmed these concerns and demonstrated that participants could discriminate between unreliable and reliable components of automated systems. Will (1991) also found that characteristics of automated agents, such as reliability, correlated with user trust in the system. Furthermore, the confidence of the user was shown to significantly impact how they interacted with the automation and the degree of trust instilled in it.

Lee and Moray (1992) reported that trust in automation does affect the operators' use of manual control if their trust is greater than their own self-confidence to perform the tasks. Riley (1994) identified self-confidence in one's manual skills as an important factor in automated usage. Riley (1989) noted that trust in the automation alone does not affect the decision to use automation, but rather a complex relationship involving trust, self-confidence, workload, skill level, and other variables determine the "reliance" factor of using automation.

Self-Efficacy and Complacency

Crew "complacency" has often been implicated as a contributing factor in aviation accidents. The term has become more prominent with the increase in automation technology in modern cockpits. As a consequence, there has also been an increase in research to understand the nature of complacency and to identify countermeasures to its onset. Parasuraman, Molloy, and Singh (1993) noted that complacency arises from overconfidence in automation reliability. They found that operators missed "automation failures" when the automated system was highly

reliable. Riley (1996) reported that an operator's decision to rely on automation may actually depend on a complex relationship between level of trust in the system, self-confidence, and other factors. Lee and Moray (1994) also found that trust in automation and self-confidence can influence decisions to use or not to use automation, but that there were large individual differences. The idea of individual differences was examined recently by Singh, Molloy, & Parasuraman (1993a). They reported a modest relationship between individual differences in complacency potential and energetic-arousal with automation-related monitoring inefficiency. Lee (1992) also conducted a number of studies examining these relationships and provided evidence that self-confidence, or self-efficacy, coupled with over estimations of automation reliability, operationally defined as trust in automation, can influence operator's decision to rely on automation.

Self-efficacy refers to expectations that people hold about their abilities to accomplish certain tasks. Bandura (1986) argued that decisions to undertake particular tasks depend upon whether or not they perceive themselves efficacious in performing those tasks. The stronger the operator's self-efficacy, the longer they will persist and exert effort to accomplish the task (Garland et al., 1988). Studies have shown that people, with higher self-efficacy for tasks, perform better in those tasks compared to people with lower self-efficacy (Bandura, 1997). However, in the aviation context, conditions can arise in which self-efficacy and the concomitant overconfidence in one's ability can impair performance. An example of this would be a pilot not off-loading tasks to automation during high workload situations because of his or her overconfidence in managing flight tasks. Therefore, the present study examined the effects of self-efficacy on automation use and complacency under high and low workload conditions.

Objectives of Present Study

The present study sought to further explore the effects of individual differences in automation use. Specifically, the generalizability of self-efficacy in monitoring performance and its relationship to automation-induced complacency was studied in addition to how the psychological construct describes the "use", "disuse", or "misuse" (Parasuraman & Riley, 1997) of automation.

Thirty participants were randomly selected from a pool of subjects who had participated in past monitoring, supervisory control, or vigilance experiments and, therefore, had a known A' score of perceptual sensitivity to critical event-to-noise ratio. Participants were equated across randomly assigned groups and then were asked to perform a 30-min vigilance task that required responses to critical events. There were 30 critical events (1 / minute) presented during the vigil and the event rate was 30 per minute. Afterwards, each participant was asked to complete task-specific self-efficacy and self-confidence questionnaires. There was no statistical difference between the participants in task performance (A'). These participants were then assigned to two experimental groups based on a median split of the self-efficacy questionnaires, which was an estimation of performance confidence in performing the task. All participants were asked to return after one week to perform a system monitoring, resource management, and tracking task from the Multiple Attribute Battery (Comstock & Arnegard, 1992) under high reliable and low reliable conditions.

The difficulty of the tasks was varied during each task run, and participants had the option to off-load the tracking task to the automation if needed. Performance, workload, and complacency measures were collected. It was hypothesized that participants who scored low in self-efficacy would be more likely to trust the automation and, therefore, exhibit complacent performance than participants who scored high in self-efficacy. However, because these

participants may be less willing to utilize automation because of the propensity to distrust automation in favor of manual skills, it was also expected that there would be a “double-edged sword” effect found. Finally, it was hypothesized that, under conditions of high mental workload, high self-efficacy participants would do significantly worse than the low self-efficacy participants who engaged automation to reduce the taskload when the task difficulty was increased.

METHOD

Participants

Thirty participants (age 18 to 39) were subjects of the study. All participants had pilot experience ranging from 0 to 110 hours of flight hours. Of the 30 participants, six had Visual Flight Rules (VFR) pilot certifications and 20 were in various stages of obtaining a private pilots license. The remaining 4 participants had a significant amount of Microsoft © Flight Simulator experience. Although flight experience was not considered an essential prerequisite for inclusion in the experiment, all participants were given a short test to assess their level of pilot knowledge. The pilots were given experimental credit toward university coursework or given the monetary amount of \$25. All participants had participated in previous vigilance studies and were randomly selected from a pool of eligible participants.

Baseline Task

The task was a 30-min simultaneous vigilance task that required participants to monitor the repetitive presentation of a pair of 3mm (W) X 38mm (H) white lines separated by 25mm. These lines appeared in the center of the monitor screen. The stimuli were white (shown black in Figure 1) and were presented on a blue background. Critical signals (targets) were 2mm taller than neutral events and occurred once a minute at random intervals. The event rate of presentation was 30 stimuli / minute. The participants were required to respond to the presence of the critical signals by pressing the space bar on the keyboard. The simultaneous vigilance task has been used in a number of studies and has been shown to be a valid task for inducing low vigilance states (Warm & Parasuraman, 1984).

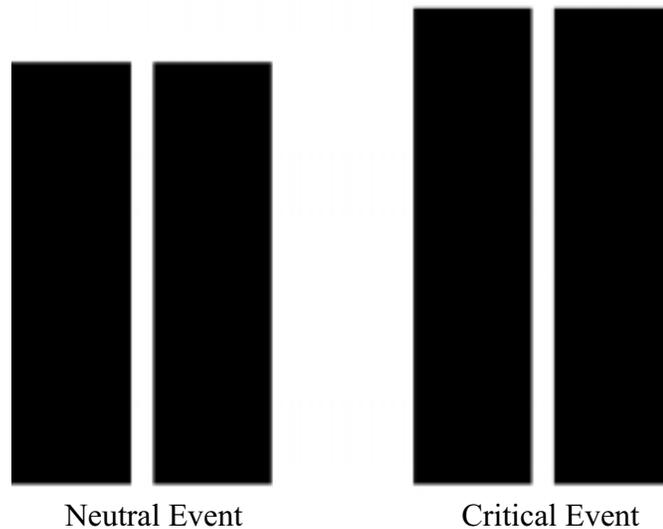


Figure 1. Baseline Vigilance Task (Not to Scale)

Experimental Task

Participants were run using a modified version of the NASA Multi-Attribute Task (MAT) battery (Comstock and Arnegard, 1992). The MAT battery is composed of four different task windows: tracking, system monitoring, communication and fuel management. These different tasks were designed to simulate the tasks that airplane crewmembers often perform during flight. Each of these tasks can be fully or partially automated. In the present study, only the tracking, monitoring, and resource management tasks were used. The monitoring task was the only task out of the three that was automated. The three tasks were displayed in separate windows of a 14-inch VGA color monitor.

Tracking Task. A two-dimensional compensatory tracking task with joystick control is presented in one window of the display (see Figure 2). The task requires participants to use the joystick to maintain a moving circle, approximately 1 cm in diameter, centered on a .5 cm by .5 cm cross located in the center of the window. Failure to control the circle results in its drifting away from the center cross. The tracking task uses a 4:3 horizontal-to-vertical sine wave driving function. The gain and difficulty levels were set at the default settings described in Comstock and Arnegard (1992).

System Monitoring. The upper-left section of the MAT battery (Figure 2) shows the system monitoring task, which consists of four vertical gauges with moving pointers and green “OK” and red “Warning” lights. Normally, the green OK light is on and the pointers fluctuate around the center of each gauge. In each 10-min block of the task, 16 “system malfunctions” occurred at unpredictable intervals ranging from 13 to 72 sec. When a system malfunction occurred, the pointer on one of the four engine gauges went “off limits”. When the engine gauge went “off limits”, the pointer shifted its center position away from the center of the vertical gauge, independent of the other 3 pointers and at intervals according to a predefined script. According to the predefined script programmed into the MAT for each task mode, the monitoring task detected 14 out of the 16 malfunctions automatically for the high reliability task mode and 9 out of the 16 malfunctions for the low reliability task mode. The red warning light came on and then went off when the automation had corrected a malfunction in 4 seconds, indicating

successful identification and correction of the malfunction. During this time, the participant's response keys were disabled to prevent manual input.

However, from time to time the automation failed to detect a malfunction. When the automation routine failed, the pointer changed its position from the center of the scale on one of the gauges independent of the other three gauges. However, the green OK light remained on and no red light appeared. The operator was responsible for detecting pointer shifts occurring on any of the four gauges, regardless of direction, and was required to respond by pressing one of the four function keys (F1, F2, F3, or F4) corresponding to the labels below each vertical gauge. Once the malfunction was detected, the pointer of the appropriate gauge moved immediately back to the center point and remained there without fluctuating for a period of 1.5 sec. (i.e. no malfunctions occurred during this time). If the participant failed to detect a malfunction, it was automatically corrected within 10 sec.

If the participant responded appropriately to an automation failure by pressing the correct function key, the response was scored as a correct detection of an automation failure. If the participant failed to detect the failure within 10 sec, the gauge was reset and the response was scored as a miss. A detection error occurred if the operator detected an automation failure but incorrectly identified the gauge associated with the failure (e.g. pressing F1 for a malfunction in engine 2). All other responses were classified as false alarms, making the performance measures for the system-monitoring task: (a) the probability of detection of automation failures, (b) reaction time (RT) for detection, and (c) the number of detection errors and false alarms made.

Fuel Management. The fuel management task is displayed in the lower, right window of the MAT batter (Figure 2). It requires participants to maintain a specific level of fuel within both of the main tanks (A & B) by selectively activating pumps to keep pace with the fuel consumption in the tanks. The six rectangular regions represent the fuel tanks. The lines that connect the tanks are pumps that can transfer fuel from one tank to another in the direction indicated by the arrow. The numbers underneath the tanks represent the amount of fuel in gallons that each tank contains. This number is updated every two seconds. The maximum amount of fuel that can be in tank A or B is 4000 gallons and in tank C or D is 2000 gallons, the remaining two tanks have unlimited capacity.

Participants were instructed to maintain fuel in tanks A and B at a tick mark that graphically depicts the level of 2500 gallons. The shaded region around the tick mark indicated acceptable performance. Tanks A and B were depleted at a rate of 800 gallons per minute and, therefore, to maintain an acceptable level of fuel, participants had to transfer fuel from one tank to another by activating one or more of the eight fuel pumps. Pressing the number key that corresponds to the pump number activates these pumps, and pressing it a second time turns it off.

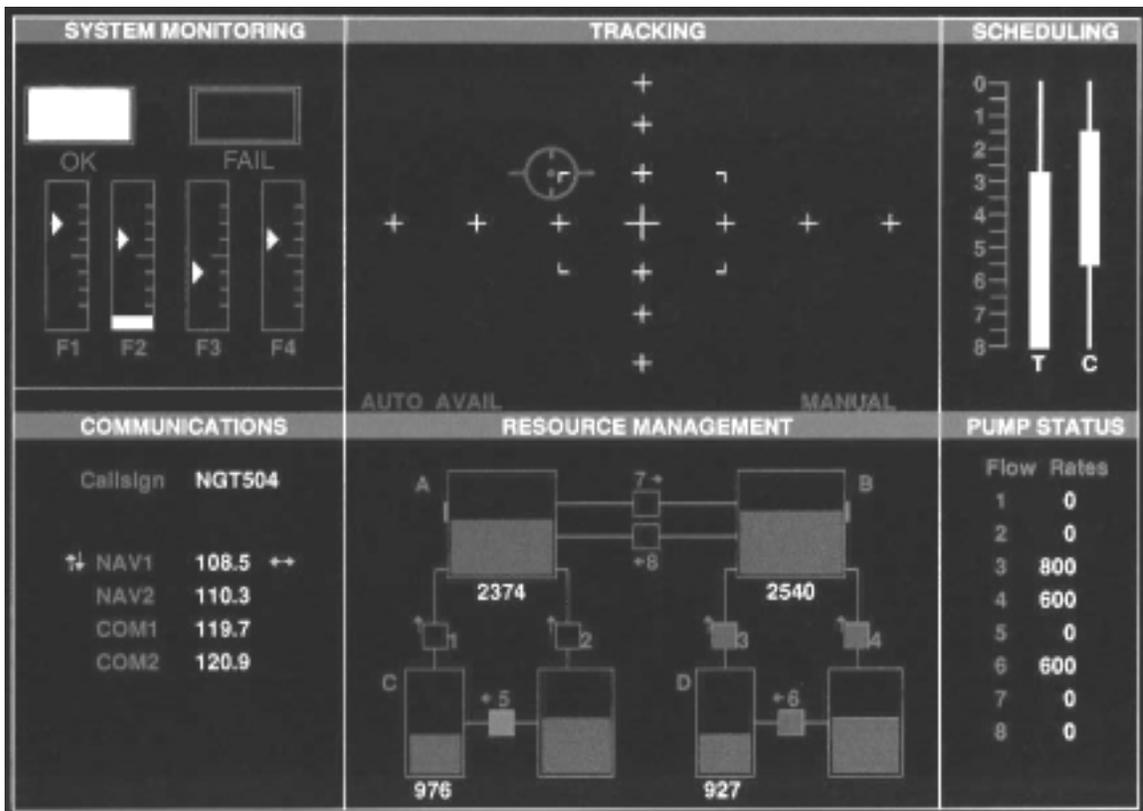


Figure 2. The Multi-Attribute Task Battery

Experimental Design

A 2 (constant, variable reliability) X 2 (high, low self-efficacy) X 2 (high, low task difficulty) mixed design was used. The independent variables were reliability of the automation condition in the system-monitoring task, self-efficacy, and difficulty. Dependent variables were A' performance, tracking RMS error, fuel level deviation, and NASA-TLX scores. The measures are discussed below under the heading of dependent measures.

The automation reliability of the system-monitoring task was defined as the percentage of 16 system malfunctions correctly detected by the automation routine in each 10-min block. The automation routine was varied as a between-subjects factor (Constant or Variable Reliability) and sessions (1-2 on consecutive days) and 10-min blocks (1-4) as within subject factors in the mixed factorial design. The reliability schedule for each condition that was employed by this study is the same one used by Singh et al. (1997). In the constant-reliability groups, the automation reliability was constant from block to block at 87.5% (14 out of 16 malfunctions detected by the automation) for each of the participants. This reliability level is used because complacency is most likely to result when working with highly reliable automated environments, in which the operator serves as a supervisory controller, monitoring system states for the occasional automation failure (Parasuraman et al., 1993). In the variable-reliability group, the automation reliability alternated every 10 min from low (9 out of 16 malfunctions detected by the automation or 56.25%) to high (87.5%) for half the participants and from high to low for the other half. No instructions about the reliability percentages of the automation were given to the participants other than the general instruction that the automation is not always reliable.

Dependent Measures

RMSE. A global measure of task performance was obtained for each participant by computing the RMS error in fuel levels of tanks A and B (deviation from the required level of 2500 gallons). Fuel levels were sampled and RMS errors computed for each 30-sec period; then they were averaged over a 10-min block to yield the RMS error for each block. Combined root-mean-square (RMS) errors were computed for samples collected over each 2-sec period and then averaged over a 10-min block to yield the mean RMS error for a given block.

A'. A' is a nonparametric measure of "perceptual sensitivity" or d' (Wickens & Hollands, 2000) and has been used extensively in the supervisory control and vigilance literature (Warm & Parasuraman, 1984). Sensitivity refers to the separation of noise and signal distributions and reflects the theoretical index of where the two normally distributed curves intersect. The value of d' is reflected by the degree to which the signal-to-noise ratio is high. The higher the ratio, the higher the value of d' than if the two curves overlay each other. Wickens and Hollands (2000) describe the hypothetical distributions and theoretical foundations of signal detection theory and the measure of d'. d' is often a theoretical value since the two curves cannot be obtained and plotted on a receiver operator characteristic (ROC). Therefore, the measure of A' can be substituted by measuring the area under the ROC curve. This provides the advantage of being parameter free and not relying on assumptions or estimations of the shape or form of the signal and noise distributions. A' can be obtained by the formula: $A' = [\text{probability (hit)} + [1 - \text{probability (false alarm)}]] / 2$. A' values range from .5 to 1 with 1 being perfect performance (i.e., perfect number of hits and no false alarms).

NASA-TLX. The NASA-TLX (Hart & Staveland, 1988) is a multi-dimensional measure of subjective workload. It requires the participant to complete a series of ratings on six 20-point scales (mental demand, physical demand, temporal demand, performance, effort, and frustration level). The "traditional" TLX scoring procedure combines the six scales, using paired comparison-derived weights, to provide a unitary index of workload. Byers, Bittner, and Hill (1989), however, demonstrated that a simple summation of responses on the six subscales produced comparable means and standard deviations, and that this "raw" procedure correlated between 0.96 to 0.98 with the paired comparison procedure. This study, therefore, summed the ratings of each scale, without the derived weighting, to provide an overall index of subjective workload for each participant.

Experimental Procedure

The thirty participants selected for participation in the present study had previous vigilance research experience. A', a measure of perceptual sensitivity (Warm & Parasuraman, 1982), measures were collected and available a priori to selection. Each participant was given a self-efficacy and self-confidence questionnaire (Bandura, 1986), which measured his or her self-perceptions of confidence in being able to complete the vigilance task. Participants were not provided any knowledge-of-results on performance.

All participants were matched for A' and questionnaire scores put into equal groups from which to randomly pool. Approximately 12 weeks later, participants were notified that they were eligible to participate in the experiment. 75% of the 40 participants agreed to be part of the present study. The thirty participants were ranked into high and low self-efficacy experimental groups based on a median split of self-efficacy questionnaire scores. A' scores were not significantly different across the groups ($p > .05$). However, because these A' scores were based on previous vigilance task performance, all participants selected for participation in the study

were asked to perform in an additional vigilance session for 30-min using the bar-type vigilance baseline task described above. Again, no differences were found in A' ($p > .05$) between the high and low self-efficacy groups.

Participants were invited back one-week later to complete four 30-minute sessions over a two-day period. Each participant received a 10-minute baseline practice session for familiarization with the MAT after having been provided detailed instructions of the functionality and operation of the task battery. After the baseline session, the experimental trials began and lasted 30-minutes each and were randomly counterbalanced for high- and low-reliability conditions. Participants were informed that the system-monitoring task was automated, and that the fuel management and tracking tasks was manual. They were informed that the automation for the system reliability task is not 100% reliable and that they were required to supervise the automation in order to respond to any malfunctions that the automation failed to detect. Participants were instructed to focus on all three tasks equally and to perform on each to the best of their ability. Participants were required to return the following day to complete the 2nd session (trials 3 and 4). There was no practice period for the second session. Two separate sessions were required because complacency has been found to be “more easily” induced under multiple sessions using a multiple-task environment (Parasuraman et al., 1993). After each experimental trial, the NASA-TLX was administered.

RESULTS

The data from the study was analyzed using a series of MANOVAs (multivariate analysis of variance) and ANOVAs (analysis of variance) statistical procedures. In all cases, alpha was set at .05 and was used to determine statistical significance. Only effects statistically significant from the MANOVA were subject to ANOVAs and are reported in the results. Expected mean squares were computed for all main effects and interactions and were used to determine error term for main effects (subjects (self-efficacy)) and interactions (reliability*subjects (self-efficacy)). Analysis of simple effects was used to examine significant interactions. Data is collapsed across experimental sessions because significant effects were not found for any dependent variable across the four 30-minute experimental sessions, $p > .05$.

A'

An ANOVA procedure was performed on the A' data for the main effects of self-efficacy (high, low) and reliability condition (variable, constant) and the interaction between self-efficacy and reliability condition. There was a main effect found for self-efficacy, $F(1, 28) = 247.82$, $p < .0001$; reliability, $F(1, 28) = 74.11$, $p < .0001$; and self-efficacy*reliability, $F(1, 28) = 39.98$, $p < .0001$. There were significant differences between high (0.89067) and low (0.73767) self-efficacy, and variable (0.856) and constant (0.77233) reliability conditions. Figure 3 graphically presents the self-efficacy*reliability interaction. Simple effects analysis determined that significant differences were between low self-efficacy, constant and all other self-efficacy, reliability combinations (3), $p < .05$.

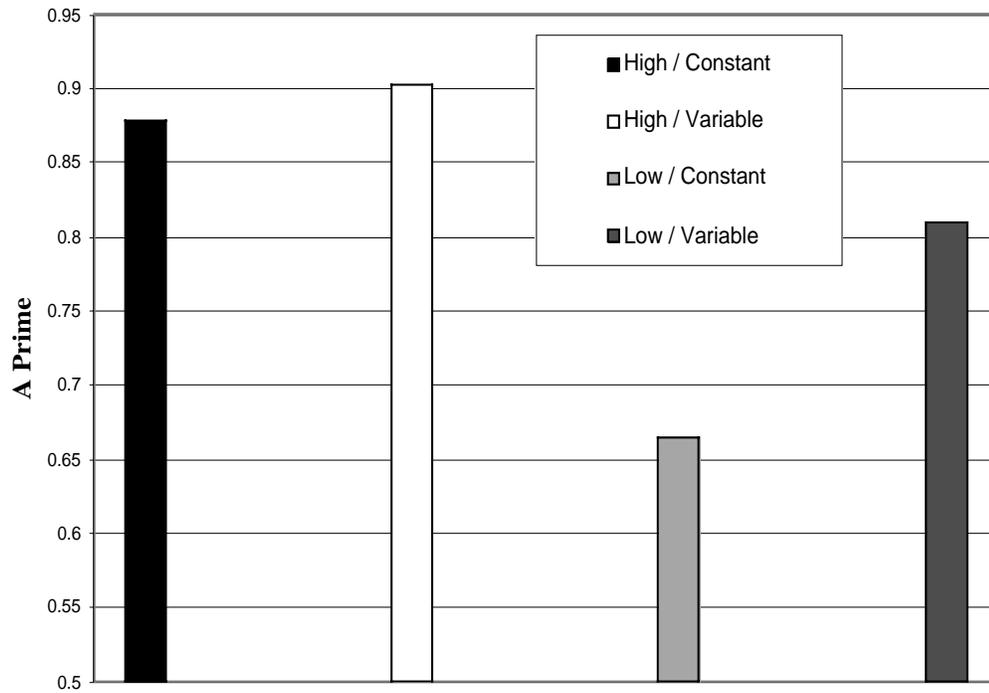


Figure 3. Reliability*Self-Efficacy Interaction for A'

RMSE

No significant differences were found for root-mean squared error for resource management tasks, $p > .05$. Overall, participants performed equally well regardless of self-efficacy (low = 48.90; high = 52.34) or reliability condition (constant = 49.34; 51.90). No further analyses were conducted on the dependent variable.

NASA Task Load Index

Byers, Bittner, & Hill (1991) showed that the NASA TLX could be analyzed using the raw scores, rather than paired comparison scaled scores, to compute workload ratings. The NASA-TLX was scored from a 0 to 100 range representing 6 subscales with 20 points. Participants rated the high workload condition (67.467) as significantly higher in workload than the low workload condition (53.200), $F(1, 28) = 115.73$, $p < .0001$, which validates the experimental manipulation of workload conditions. An ANOVA also revealed that, overall, high self-efficacy participants (62.667) rated the tasks as significantly higher in workload than low self-efficacy participants (58.00), $F(1,28) = 12.38$, $p < .05$.

The main effect for self-efficacy TLX ratings must be considered in perspective of the interaction effect found between self-efficacy and reliability, $F(1,28) = 1120.90$, $p < .0001$, and presented in Figure 4. A simple effects analysis found that high self-efficacy participants, under the high workload condition, rated the task as significantly higher in workload than the other 3 combinations of self-efficacy and reliability (i.e., low self-efficacy / low workload; low self-

efficacy, high workload; high self-efficacy / low workload). Moreover, low self-efficacy participants rated the task significantly higher under the LOW workload condition than the high workload condition probably due to the off-loading of the tracking task to the automation during the high workload condition. High self-efficacy (100%) did not take advantage of the option to off-load the automation during this time whereas 87% of low self-efficacy participants did.

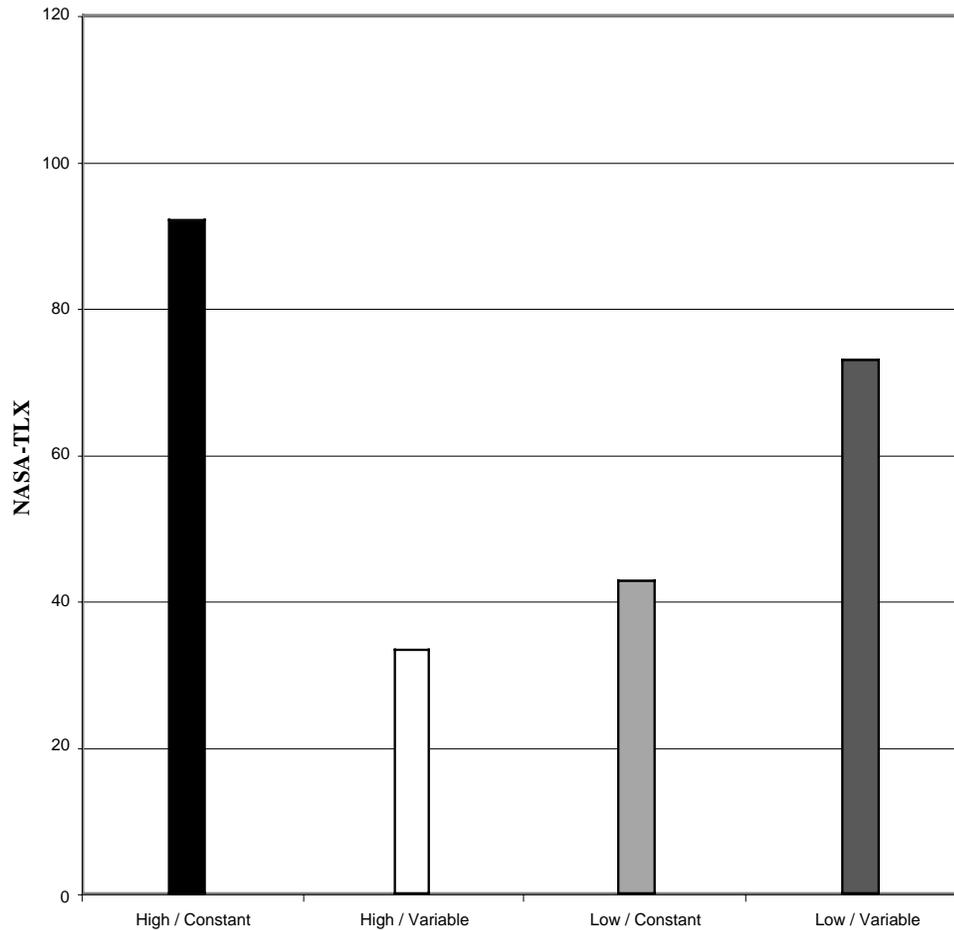


Figure 4. Reliability*Self-Efficacy Interaction for NASA-TLX

DISCUSSION

As predicted, participants with high self-efficacy did significantly better in both the constant and variable reliability conditions. Only participants with low self-efficacy were found to have suffered automation-induced complacency. However, these participants did significantly better than high self-efficacy participants during the high workload condition. Parasuraman, Molloy, and Singh (1993) reported that complacency arises only in multiple task situations with concomitant increases in workload. Therefore, offloading the task reduced the workload demands significantly enough, and may have freed up cognitive resources that allowed low self-efficacy participants to perform the systems monitoring task more effectively than high self-efficacy participants. These participants did not take the option of off-loading the tracking task to

the automation when given opportunity to do so. This is confirmed by the fact that low self-efficacy participants rated workload to be significantly lower in the high workload condition than high self-efficacy participants.

The results of the study suggest that self-efficacy is an important moderator variable of whether an operator will succumb to automation-induced complacency. Having low self-efficacy, a reflection of one's perception of ability, regardless of skill level, can set-up cognitive strategies that may increase the potential for succumbing to automation-induced complacency. However, high self-efficacy may serve as a double-edged sword in producing overconfidence in one's ability that may limit other strategies, such as task offloading, for managing cognitive workload. Future research should extend these findings to actual pilot populations to determine the relationship of self-efficacy to automation-induced complacency and how self-efficacy can moderate the pilot-automation interaction in supervisory control environments.

REFERENCES

- Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall
- Bandura, A. (1997). Self-efficacy in changing societies. Englewood Cliffs, NJ: Prentice-Hall.
- Barber, B. (1983). Logic and the limits of trust. New Brunswick, NJ: Rutgers University Press.
- Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1991). Effects of feedback on workload in visual performance. In Proceedings of the Human Factors and Ergonomics Society 35th Annual Meeting (pp. 1491-1494). Santa Monica, CA: Human Factors and Ergonomics Society.
- Billings, C. E., Lauber, J. K., Funkhouser, H., Lyman, G., & Huff, E. M. (1976). NASA Aviation Safety Reporting System (Tech. Rep. TM-X-3445). Moffett Field, CA: NASA Ames Research Center.
- Billings, C. E. (1997). Aviation automation: The search for a human-centered approach. New Jersey: Lawrence Erlbaum Associates.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology, 21, 1-16.
- Byers, J.C., Bittner, A.C., & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), Advances in Industrial Ergonomics and Safety I. London: Taylor & Francis.
- Byrne, E.A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological Psychology, 42, 249-268.
- Comstock, J. R., & Arnegard, R. J. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research (Tech. Memorandum No. 104174). Hampton, VA: NASA Langley Research Center.
- Damos, D.L., Smist, T.E., & Bittner, A.C. (1983). Individual differences in multiple-task performance as a function of response strategy. Human Factors, 25, 215-226.
- Deaton, J. E., & Parasuraman, R. (1993). Sensory and cognitive vigilance: Effects of age on performance and subjective workload. Human Performance, 6, 71-97.
- Dittmar, M. L., Warm, J. S., Dember, W. N., & Ricks, D. F. (1993). Sex differences in vigilance performance and perceived workload. Journal of General Psychology.
- Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 41-62). North-Holland: Elsevier Science Publishers.
- Farmer, R., & Sundberg, N. D. (1986). Boredom proneness- the development and correlates of a new scale. Journal of Personality Assessment, 50, 4-17.
- Fulop, A., & Scerbo, M. W. (1991, September). The effects of event rate on perceived workload and boredom in vigilance. Poster presented at the Human Factors Society 35th Annual Meeting, San Francisco, CA.

- Galinsky, T. L., Dember, W. N., & Warm, J. S. (1989, March). Effects of event rate on subjective workload in vigilance. Poster presented at the Human Factors Society 35th Annual Meeting, San Francisco, CA.
- Grubb, P. L., Miller, L. C., Nelson, W. T., Warm, J. S., & Dember, W. N. (1994). Cognitive Failure and Perceived Workload in Vigilance Performance. Proceedings from the First Automation Human Performance Conference (p. 115-121) Washington, D.C.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.
- Langer, E. (1989). Mindfulness. Reading, MA: Addison-Wesley.
- Lee, J. D. (1992). Trust, self-confidence, and adaptation to automation. Unpublished Doctoral Thesis. University of Illinois, Urbana, IL.
- Lee, J. D., & Moray, N. (1992). Trust and the allocation of function in the control of automatic systems. Ergonomics, *35*, 1243-1270.
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, *40*, 153-184.
- Mosier, K.L., Skitka, L.J., & Korte, K.J. (1994). Cognitive and social psychological issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: current research and trends (pp. 256-263). Hillsdale, NJ: Earlbaum.
- 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 and use of automation controllers in a supervisory process control task. Doctoral dissertation, University of Toronto, Canada.
- Muir, B. M. (1994). Trust in automation. Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics, *37*, 1905-1922.
- Muir, B. M. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, *39* (3), 429-460.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". International Journal of Aviation Psychology, *3* (1), 1-23.
- Parsons, H.M. (1985). Automation and the individual: Comprehensive and comparative views. Human Factors, *27*, 99-112.
- Pope, A. T. & Bogart, E. H. (1992). Identification of hazardous awareness states in monitoring environments. SAE Technical Paper #9211369, SAE 1992 Transactions: Journal of Aerospace, Sec. 1-REM 101 pp.449-457.
- Prinzel, L. J. & Freeman, F. G. (1997). Task-specific sex differences in vigilance performance. Subjective workload and boredom. Perceptual & Motor Skills, *85* (3), 1195-1202.

- Prinzel, L. J., Pope, A. T., & Freeman, F. G. (1999). The double-edged sword of self-efficacy: Implications in automation-induced complacency. In Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting. (pp. 434). San Diego, CA: Human Factors and Ergonomics Society.
- Riley, V. (1989). A general model of mixed-initiative human-machine systems. In Proceedings of the Human Factors and Ergonomics Society 33rd Annual Meeting, (pp. 124-128). Denver, CO: Human Factors and Ergonomics Society.
- Riley, V. (1994). A theory of operator reliance of automation. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 8-14). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), Automation Theory and Applications. Human Factors in Transportation, (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sawin, D. A., & Scerbo, M. W. (1994). Vigilance: How to do it and who should do it. In Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting (pp. 1312-1316). Santa Monica, CA: Human Factors and Ergonomics Society.
- Sawin, D. A., & Scerbo, M. W. (1995). Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload. Human Factors, *37* (4), 752-765.
- Scerbo, M. W., Greenwald, C. Q., & Sawin, D. A. (1993). The effects of subject-controlled pacing and task type upon sustained attention and subjective workload. Journal of General Psychology, *113*, 293-307.
- Scerbo, M. W., Rettig, K. M., & Bubb-Lewis, C. L. (1994). A validation study of a task-related boredom scale. In Proceedings of the 2nd Mid-Atlantic Human Factors Conference, (pp. 135-136). Washington, D. C.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1993a). Automation-induced "complacency": Development of a complacency-potential scale. International Journal of Aviation Psychology, *3*, 111-122.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1993b). Individual differences in monitoring failures of automation. Journal of General Psychology, *120* (3), 357-373.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. International Journal of Human-Computer Studies, *46*, 17-30.
- Stark, J. M. & Scerbo, M. W. (1998). The effects of complacency potential and boredom proneness on perceived workload and task performance in an automated environment. Presented at the Human Factors and Ergonomics Society 42nd Annual Meeting. Chicago, IL: Human Factors and Ergonomics Society.
- Thackray, R. I., & Touchstone, R. M. (1989). Detection efficiency on an air traffic task with and without computer aiding. Aviation, Space, and Environmental Medicine, *60*, 744-748.
- Warm, J. & Parasuraman, R. (1984). The psychology of vigilance. Mahwah, NJ: Lawrence Erlbaum Associates.

- Wicken, C.D., & Hollands, J.G. (2000). Engineering psychology and human performance. Upper Saddle River, NJ: Prentice Hall.
- Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: Promises and problems. Ergonomics, *23*, 995-1011.
- 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. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft (Contractor Report 117528). Moffett Field, CA: NASA-Ames Research Center.
- Will, R. P. (1991). True and false dependence on technology: Evaluation with an expert system. Computers in human behavior, *7* (3), 171.
- Wise, J. & Hopkins, D. (2000). Human Factors in Certification. Mahwah, NJ: Lawrence Erlbaum Associates.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 2002	3. REPORT TYPE AND DATES COVERED Technical Memorandum	
4. TITLE AND SUBTITLE The Relationship of Self-Efficacy and Complacency in Pilot-Automation Interaction			5. FUNDING NUMBERS 711-50-21-01	
6. AUTHOR(S) Lawrence J. Prinzel III				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199			8. PERFORMING ORGANIZATION REPORT NUMBER L-18231	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA/TM-2002-211925	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified-Unlimited Subject Category 54 Distribution: Standard Availability: NASA CASI (301) 621-0390			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Pilot "complacency" has been implicated as a contributing factor in numerous aviation accidents and incidents. The term has become more prominent with the increase in automation technology in modern cockpits and, therefore, research has been focused on understanding the factors that may mitigate its effect on pilot-automation interaction. The study examined self-efficacy of supervisory monitoring and the relationship between complacency on strategy of pilot use of automation for workload management under automation schedules that produce the potential for complacency. The results showed that self-efficacy can be a "double-edged" sword in reducing potential for automation-induced complacency but limiting workload management strategies and increasing other hazardous states of awareness.				
14. SUBJECT TERMS Automation, Complacency, Self-Efficacy, Workload, Selection			15. NUMBER OF PAGES 26	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	