NASA/TM-2001-211413

Examination of Automation-Induced Complacency and Individual Difference Variates

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

Holly DeVries, Fred G. Freeman, and Peter Mikulka
Old Dominion University, Norfolk, Virginia

December 2001

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-2001-211413

Examination of Automation-Induced Complacency and Individual Difference Variates

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

Holly DeVries, Fred G. Freeman, and Peter Mikulka
Old Dominion University, Norfolk, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

December 2001

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

Automation-induced complacency has been documented as a cause or contributing factor in many airplane accidents throughout the last two decades. It is surmised that the condition results when a crew is working in highly reliable automated environments in which they serve as supervisory controllers monitoring system states for occasional automation failures. Although many reports have discussed the dangers of complacency, little empirical research has been produced to substantiate its harmful effects on performance as well as what factors produce complacency. There have been some suggestions, however, that individual characteristics could serve as possible predictors of performance in automated systems. The present study examined relationship between the individual differences of complacency potential, boredom proneness, and cognitive failure, automation-induced complacency. Workload and boredom scores were also collected and analyzed in relation to the three individual differences. The results of the study demonstrated that there are personality individual differences that are related to whether an individual will succumb to automation-induced complacency. Theoretical and practical implications are discussed.

TABLE OF CONTENTS

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, however, do not escape visibility so easily, such as the modern airplane. Wiener and Curry (1980), and Wiener (1989) noted that avionics has provided a dramatic increase in airline capacity and productivity coupled with a decrease in manual workload and fatigue, more precise handling as well as relief from certain routine operations, and more economical use of airplanes. But, unlike the washing machine, the increase in automation in airplanes and air navigational systems, 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 an increase in the number of errors caused by pilot-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 ½ mile 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). Pilots originally manually performed the tasks of extending the flaps and slats, the maneuvering needed if a stall does occur with the airplane, and 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 take-off and failed to confirm the configuration with the use of the taxi checklist (Billings, 1997).

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 is capable of detecting and responding 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 of both studies, may be a result of the participant's premature cognitive commitment or perceived trust in the automation to correct for system failures.

Trust

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 and in this role trust in automation affected pilot-automation 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.

Lee (1992) conducted a number of studies examining these relationships and provided evidence that self-confidence coupled with trust influence operator's decision to rely on automation. Prinzel, Pope, and Freeman (1999) found that participants with high self-confidence and manual skills did significantly better with a constant, highly reliable automated task, than participants who had lower confidence in their own monitoring ability. The high self-confidence participants also rated workload significantly higher suggesting a micro-tradeoff; participants were able to maintain monitoring efficiency but at the cost of higher workload. Participants with lower self-confidence and manual skills, however, did significantly poorer in monitoring the automated task under the reliable automation condition suggesting the onset of complacency.

Assessment of Complacency

Singh, Molloy, and Parasuraman (1993b) noted that complacent behavior may often coexist with other conditions. Examples include the following: (a) operator experience with equipment; (b) high workload; and (c) fatigue due to poor sleep or exhaustion. They state that "...the combination of the crew's attitude toward automation (e.g. overconfidence) and a particular situation (e.g. high workload) may lead to complacent behavior." Therefore, pilot attitudes of overconfidence and over-reliance on automation may not, alone, produce conditions of complacency, but instead may indicate a potential for complacency. These authors developed a 20-item Complacency-Potential Rating Scale (CPRS) for measuring such attitudes toward general automated devices, such as automated teller machines (ATMs) and VCRs. A factor analysis of the CPRS indicated that the major factors that contribute to a person's "complacency potential" were trust in, reliance on, and confidence in automation. Singh and his colleagues (1993a) further demonstrated that complacency potential was not correlated with the constructs measured on either the Eysenck Personality Inventory or the Thayer Activation-Deactivation Adjective Check List, scales often used in vigilance and monitoring research, suggesting their relative independence.

Although, the CPRS has been shown to be a good indicator of an operator's complacency potential it is not able to discriminate between the number of possible factors involved in the occurrence of automation-induced complacency. The scale does not measure the other factors that may influence the onset of complacency such as workload, boredom, or cognitive failure. Therefore, other measures are also needed to assess it fully.

Parasuraman et al. (1993) demonstrated that the performance measures: (a) the probability of detection of automation failures, (b) reaction time (RT) to detection, and (c) the number of false alarms and detection errors made could be used to assess the consequences of complacency in a multi-task environment. Subjective scales of various psychological constructs may also be valuable tools when assessing automation-induced complacency.

Boredom and Workload

Mental workload refers to the amount of processing capacity that is expended during task performance (Eggemeier, 1988). Riley (1996) noted that although workload was a necessary aspect of automation-induced complacency, little workload-related research exists.

Parasuraman and his colleagues (1993), found the low workload level of a single task condition, consisting of only a system-monitoring task, was not sufficient to induce complacency. They reasoned that in a single-task environment a state of boredom would be experienced by the subjects, due to the low workload level involved in the task. The detection rates, however, for both high and low reliability groups in this condition were extremely high (near 100%). Therefore, they concluded that the lack of complacency experienced by participants in the single-task condition suggested that complacency and boredom are two distinct concepts.

In contrast, several studies have linked boredom, especially the propensity to become bored, to high amounts of workload. Sawin and Scerbo (1994, 1995) in their use of vigilance tasks report that boredom often has a high workload aspect associated with it. The information-processing demands or workload experienced by participants performing a vigilance task were once thought to be minimal. Fulop and Scerbo (1991), however, have recently demonstrated that participants find vigilance tasks to be stressful and other researchers have found them to be demanding due to the high workload involved in remaining vigilant (Deaton & Parasuraman, 1993; Galinsky, Dember, & Warm, 1989).

Farmer and Sundberg (1986) isolated a single measurable trait, *boredom proneness* (BP), which they report as highly related to a person’s tendency to become bored. They developed a 28-item scale, the Boredom Proneness Scale (BPS: Farmer & Sundberg, 1986), to measure this trait. Stark and Scerbo (1998) found significant correlations between workload, complacency potential, and boredom proneness, by examining their effects on task performance using the Multi-Attribute Task Battery (MAT; Comstock & Arnegard, 1992). Their study supports the view that the psychological state of boredom may be a factor that induces complacency. The results of Parasuraman et al. (1993) thus need to be considered cautiously since they reported no workload or boredom data to support their claim that their single task represented an underloaded task condition that caused boredom and, therefore, that boredom and complacency are unrelated. A considerable amount of evidence points to high workload being associated with boredom components while performing supervisory control and vigilance tasks (Becker, Warm, Dember, & Hancock, 1991; Dittmar, Warm, Dember, &

Ricks, 1993; Prinzel & Freeman, 1997; Scerbo, Greenwald, & Sawin, 1993). In addition, Pope and Bogart (1992) reported that ASRS reports contain descriptions of crews becoming “complacent” due to succumbing to “boredom” and “experiences of diminishing attention, compromised vigilance, and lapsing attention, frequently not associated with fatigue” (p. 449). Therefore, automation-induced complacency may be composed of a number of dimensions including trust, boredom proneness, complacency potential, self-confidence, skill-level, workload management ability, and experience to name a few. All of these dimensions are or can be influenced by the individual differences of each human operator. For example, Riley (1989) stated that trust is a multidimensional construct that has both cognitive and emotive qualities that can be influenced by individual differences.

Grubb, Miller, Nelson, Warm, and Dember (1994) examined one such personality dimension, “cognitive failure” and its relation to perceived workload in vigilance tasks, as measured by the NASA-TLX. They reported that operators high in cognitive failure (HCF) tend to be more absent-minded, forgetful, error-prone, and less able to allocate mental resources to perform monitoring tasks than those classified as low in cognitive failure (LCF; Broadbent, Cooper, Fitzgerald, & Parkes, 1982). Interestingly, Grubb et al. (1994) found HCF and LCF participants performed equally well on vigilance tasks but the workload scores of the HCF were significantly higher than their LCF peers; thus, these participants performed as well as LCF participants but did so at a higher cost in resource expenditure. The HCF individuals, therefore may exhibit complacent behaviors, due to their resources being largely depleted, when faced with continuing a task. This prevalence towards cognitive failure may be another factor related to a person’s becoming complacent while monitoring automation.

The individual differences described above suggest that automation-induced complacency may represent a complex dynamic of many psychological constructs. As Singh et al. (1993) describe, “...the psychological dimensions of complacency and its relation to characteristics of automation are only beginning to be understood....” and that other individual and social factors may also play a role. Therefore, a need remains to examine other psychological antecedents that may contribute to automation-induced complacency.

Present Study

The present study is an exploratory examination of automation-induced complacency in relation to the personal dimensions of: complacency potential, boredom proneness, and cognitive failure. All of these dimensions are hypothesized to have an effect on whether an individual will experience complacency within a multi-task environment. “Complacency will be defined as the operator failing to detect a failure in the automated control of the system monitoring task,” (Parasuraman et. al, 1993 p. 4). The conditions likely to lead to poor monitoring of automation will be manipulated by having the reliability of the system-monitoring task remain constant or variable over time (Parasuraman et al., 1993; Singh et al., 1997).

Automation-induced complacency is a complex psychological construct, which may be influenced by the individual differences of the human operator. Therefore, the relationship of the individual differences, workload, and boredom scores to the efficiency of monitoring for automation failures will be examined.

Research Hypotheses

1. A partial replication of the Singh et al.'s (1997) methods were performed using constant and variable reliability of the system-monitoring task automation. In the constant-reliability group, the automation reliability was constantly high at 87.5% (14 out of 16 malfunctions detected, and in the variable-reliability group, it alternated every 10 min from low (56.25%) to high (87.5%) for half the participants and high to low for the other half. Participants in the constant-reliability condition were hypothesized to experience complacency, indicated by low performance on the system-monitoring task, relative to participants in the variable reliability condition.

2. The Complacency Potential Rating Scale (CPRS; Singh et al., 1993) measures attitudes toward automation that reflect a potential for developing automation-induced complacency. Participants who scored high on the CPRS were hypothesized to perform significantly worse on the system-monitoring task than participants who were low in complacency potential, in the constant reliability condition. No differences were expected between the two groups in the variable reliability condition.

3. The constant-reliability condition has a lower automation failure rate, which allowed participants to peripheralize the system-monitoring task, as they trust the automation to fix any malfunction. Therefore, in the current study participants in the constant-reliability condition are expected to perform the tracking task significantly better than participants in the variable-reliability condition.

4. The resource management task has been shown to require few cognitive resources to perform it adequately (i.e. keep the fuel tanks at approximately 2500 gallons). No significant differences where expected between the constant-reliability group's performance and the variable-reliability group's performance.

5. Participants who are classified as high complacency potential were expected to rate the task-related boredom of the MAT-Battery to be significantly higher than those participants who are low in complacency potential.

6. The NASA-TLX was used to assess the participants' subjective workload for each task condition. No significant differences were expected between the groups because all participants were required to perform all three of the MAT-battery tasks.

7. The Complacency Potential Rating Scale, Boredom Proneness Scale, and the Cognitive Failure Questionnaire have all been used in previous studies to examine individual differences of human behavior. In the current study the three scales were used to examine the individual differences of each participant because each is hypothesized to have an effect on whether a participant will experience automation induced complacency.

Therefore, a significant correlation was expected between the three individual difference scales.

8. As discussed previously, the personality dimension “cognitive failure” may be a precursor to participants becoming complacent and result in poor performance when monitoring automation. A significant negative correlation was expected between scores on the Cognitive Failure Questionnaire and performance on the system- monitoring task.

9. Individuals who are high in cognitive failure (HCF) experience a higher cost in resource expenditure when performing multiple tasks than low cognitive failure (LCF) individuals (Grubb et al., 1994). A significant positive correlation was expected between scores on the CFQ and workload scores on the NASA-TLX.

10. The present study was also interested in how a person’s level of susceptibility to boredom may contribute to automation induced complacency. As with the previous hypotheses, those concerning boredom proneness are exploratory in nature as little research exists on how boredom affects complacency behavior. The Boredom Proneness Scale (BPS; Farmer and Sundberg, 1986) was used as used to measure boredom as a trait and assess each individual’s proneness to become bored. A significant negative correlation was hypothesized to occur between scores on the Boredom Proneness Scale and performance on the system-monitoring task.

11. A positive correlation was expected between scores on BPS and scores on the NASA-TLX, which demonstrates perceived workload.

METHOD

Participants

Forty undergraduate students from Old Dominion University received extra credit or $20.00 for their participation in this study. The experimental design of the study was approved by the Old Dominion University Internal Review Board for the use of human participants, prior to participation recruitment. The ages of the participants were 18 to 40. All participants completed the study voluntarily and all had normal (20/20) or corrected-to-normal vision.

Experimental Design

The three individual difference measures, Complacency-Potential Rating Scale (CPRS; Singh et al., 1993), Cognitive Failure Questionnaire (CFQ; Broadbent et al., 1982), and the Boredom Proneness Scale (BPS; Farmer and Sundberg, 1986) were used to measure these traits in each participant. The NASA-TLX (task-load index; Hart & Staveland, 1988) and the Task-related Boredom Scale (TBS; Scerbo et a., 1994) were used to assess the total subjective workload and total perceived boredom experienced by each participant, respectively.

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.

Participants were classified as either high complacency or low complacency based on their score on the Complacency-Potential Rating Scale. A median split procedure was used for this classification as recommended by Singh et al. (1993). Singh et al. (1993) used a median split of 56 to classify their participants. The median split for the current study was 58, which was computed after the first ten and then the first twenty participants had completed the experiment, and thus was used to classify participants as low or high complacency. Fifty-eight was also the median split once all 40 participants had completed the study. Once classified as high or low complacency the participants were randomly assigned to one of the two experimental (automation-reliability) conditions.

This grouping process resulted in an equal number of high and low complacency participants being placed in each experimental condition.

A multivariate analysis of variances (MANOVA) was calculated on all the performance data collected. Separate analysis of variances (ANOVAs) were computed for each task on the MAT battery (fuel management, tracking, and system monitoring), with complacency potential as the sub-grouping variable. A 2 (constant or variable automation reliability) X 2 (sessions) X 4 (10 min blocks) X 2 (median split of CPRS) mixed factorial design was employed for these analyses.

Experimental Tasks

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 1). 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). 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.

System Monitoring. The upper-left section of the MAT battery (Figure 1) 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 1). 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.

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.

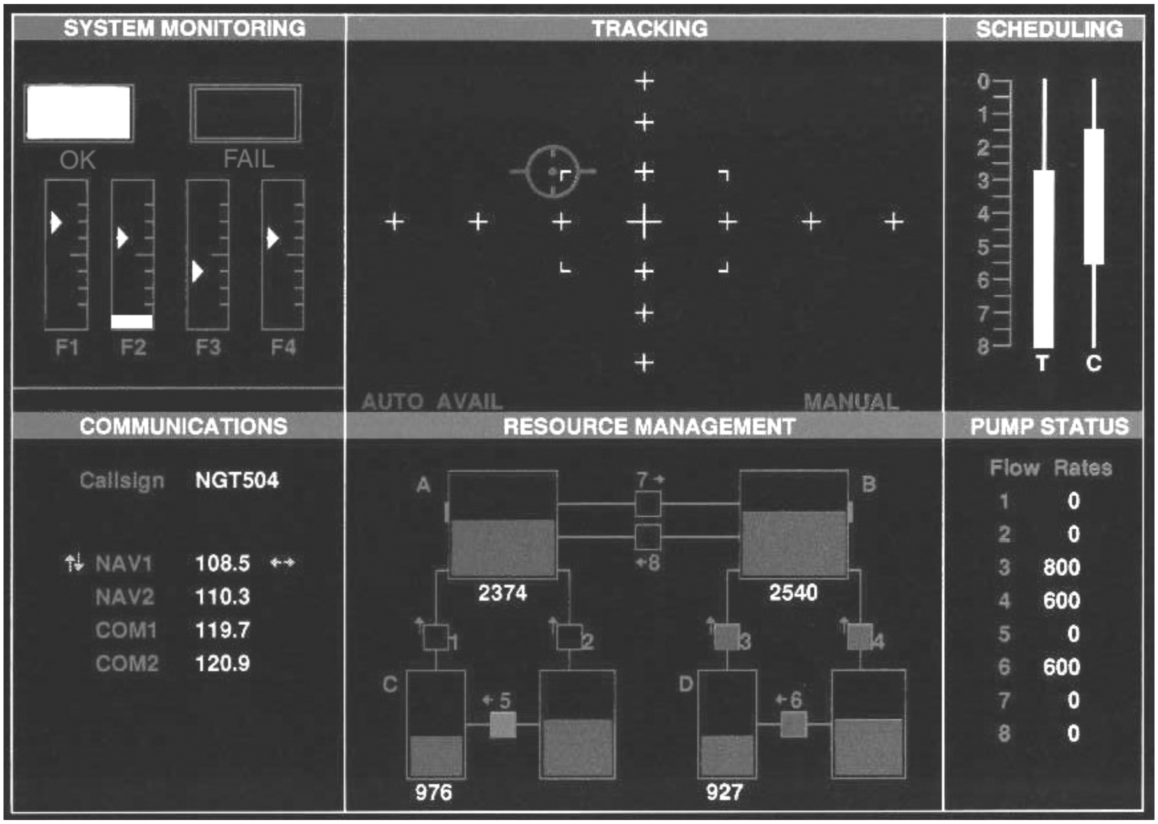


Figure 1. The Multi-Attribute Task Battery

Individual Difference Measures

Median Split Procedure. Participants were classified as low-complacency potential or high-complacency potential on the basis of a median score split obtained on the measure. The developers of the Complacency Potential Rating Scale (CPRS; Singh et al. 1993) recommend this grouping procedure which they utilized in their research on individual differences using the CPRS. The median split of 58 was computed for the current study after the first ten and the first twenty participant completed the experiment, and thus was used to classify participants as low or high complacency potential.

Complacency Potential Rating Scale (CPRS). The CPRS was developed to measure people's attitudes toward general automated devices, such as automatic teller machines (ATMs) and VCRs. Factor analysis by Singh et al. (1993a) of sample responses (N = 139) of the scale indicated four complacency-potential related dimensions: trust, confidence, reliance, and safety. Singh and colleagues suggest that high scores on these factors are associated with complacency.

The CPRS has 20 items, including 4 filler items. Each item has a scale ranging from *strongly disagree* (1) to *strongly agree* (5). The range of scores on the CPRS can vary from 16 (low complacency potential) to 80 (high complacency potential). The measure has high internal consistency ($r > .98$) and test-retest reliability ($r = .90$) among the items. (Singh et al., 1993). The total CPRS score for each participant is obtained by adding scores for items 1-16, excluding the scores for the 4 filler statements. The participants are then classified as either high- or low-complacency potential on the basis of a median split of the CPRS scores. (See Appendix A).

Boredom Proneness Scale (BPS). The BPS was developed by Farmer and Sundberg (1986), as a general assessment tool to measure the tendency to experience boredom. According to Farmer and Sundberg (1986), " boredom is a common emotion, with boredom proneness a predisposition with important individual difference, (p.4)."

The current 28-item dichotomous self-report scale asks participants to answer "yes" or "no" to each item. Items include statements such as "It is easy for me to concentrate on my activities" and "It takes more stimulation to get me going than most people." Farmer and Sundberg (1986) report an acceptable internal reliability ($\alpha = .79$) and test-retest reliability ($r = .83$). See Appendix B. The BPS has also demonstrated predictive validity in the evaluation of interest and attention in the classroom and has been shown to be correlated with other measures of boredom, such as Lee's Job Boredom Scale ($r = .49$, $p < .001$), (Lee, 1983). See Appendix B.

Cognitive Failures Questionnaire (CFQ). Broadbent et al. (1982), developed the CFQ, as a self-report inventory which measures failures in perception, memory, and motor function. Participants who score as high in cognitive failure (HCF) tend to be more absent-minded, forgetful, error-prone, and less able to allocate mental resources to perform monitoring tasks than those classified as low in cognitive failure (Grubb et al., 1994).

The scale consists of 25 items to evaluate lapses in attention, slips of action, and failure of everyday memory. The items include such statements as "Do you read something and find you haven't been thinking about it and must read it again?" and "Do you fail to notice sign posts on the road?" Participants are to indicate whether these minor mistakes have happened to them in the last six months: very often, quite often, occasionally, very rarely or never. Each item has a 0-4-point value and a participant's score on the scale range can range from 0-100. Broadbent et al. (1982) report an acceptable level of internal consistency ($\alpha = .89$) and test-retest reliability ($r = .82$). See Appendix C.

NASA Task Load Index (TLX; Hart & Staveland, 1988). The NASA-TLX 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, combined the ratings of each scale to provide an overall index of subjective workload for each participant. See Appendix D.

Task-related Boredom Scale (TBS; Scerbo, Rettig, & Bubb-Lewis, 1994). The TBS addresses eight factors thought to contribute to feelings of boredom: stress, irritation, relaxation, sleepiness, alertness, concentration, passage of time, and satiation. In addition, respondents are also asked to provide an estimation of their overall feeling of boredom. A total boredom score is calculated by summing all the subscales. The sleepiness, time passage and desire for task to end are reversed scored. See Appendix E.

Procedure

Upon entering the laboratory each participant was given the Complacency-Potential Rating Scale (CPRS) to complete. Once the participant had completed the CPRS and returned it to the experimenter, he or she was given the Boredom Proneness Scale and the Cognitive Failure Questionnaire to fill out. While the participants completed the remaining two individual difference measures, their CPRS scores were calculated and they were classified as high-complacency potential or low-complacency potential based initially on the median-split score of 56 recommended by Singh et al. (1993). However, after the first ten participants had completed the study a median split for this experiment was 58, which continued to be the case when computed a second time after twenty participants had completed the study. The new median split was then used and all participants were re-classified according to the new median split. The new split affected only a few of the participants, reclassifying them as low or high complacency respectively. The median split was also computed once all forty participants completed the study, and it remained 58. Once each participant had been sub-grouped as either high or low complacency potential, and completed all the pre-experimental measures, he or she was randomly assigned to one of the two experimental (automation-reliability) conditions (i.e. constant or variable reliability), with the restriction that an equal number of high and low complacency participants were placed in each experimental group.

Participants were tested individually, completing two 40-min computer sessions over a period of 2 days (one session per day). Each participant was instructed individually on the components of the MAT battery and given a 10-min practice session in which they performed all three tasks manually. Each participant was asked to give equal attention to each of the three tasks. After a 3-min rest period, the experimental session began. Participants were informed that the system-monitoring task was automated, and that the fuel management and tracking tasks were 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. At the end of each session, each participant was required to complete the NASA-TLX and TBS. Participants were required to return the following day or as soon as possible to complete the 2nd session. 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). The NASA-TLX and TBS were filled out again after the completion of the second session. The appropriate paperwork for receiving the 3 extra credits or $20 was filled out before the participant left on the second day of testing.

RESULTS

Complacency is a complex psychological phenomenon, which has yet to be fully defined, and there may be many different variables that are involved in why some individuals experience it and others do not. Therefore, this study also examined whether people who experience automation-induced complacency also tend to score high in cognitive failure and boredom proneness and experience high workload and high amounts of task related boredom.

The primary independent variables of the study included Reliability Condition (RC), and the level of Complacency Potential (CP). The individual difference variables analyzed were the scores on the Boredom Proneness Scale (BPS), and the Cognitive Failure Questionnaire (CFQ). The two reliability condition groups included constant high 87.5% accuracy or variable reliability, which fluctuated from high (87.5%) to low (56.25%) reliability every ten-minute block. The median split of 58 on the Complacency Potential Rating Scale designated the high or low complacency potential groups before each participant began the study. The dependent variables included scores on NASA-TLX and Task-related Boredom Scale, tracking (RMSE), resource management performance (deviations from standard criteria of 2500 gallons in tanks A and B), and system monitoring performance (A').

A correlational analysis was performed on the measures, Cognitive Failure Questionnaire, Boredom Proneness Scale, the Complacency Potential Rating Scale, the NASA-TLX and the Task-Related Boredom Scale for each formal hypothesis. The performance 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 level was set at .05 and was used to determine statistical significance. An analysis of simple effects was used to examine all significant interaction effects.

Task Performance

A MANOVA was analyzed for the performance tasks variables: tracking, resource management and system-monitoring, on the NASA Multi-Attribute Task (MAT) battery (Comstock and Arnegard, 1992). A significant main effects were found for reliability condition, $F(8,29) = 7.6102$, $\underline{p} < .0001$ and complacency potential , $F(8,29) = 37.4148$, $\underline{p} < .0001$. It also demonstrated a significant interaction of complacency potential and reliability condition, $F(8,29) = 7.5959$, $\underline{p} < .0001$, See Table 1. Subsequent ANOVA procedures were then performed for each of these significant effects.

Table 1
Multivariate Analysis of Variance

	Value	F	Num DF	Den DF
RC Effect	0.32264590	7.6102	8	29
CP Effect	0.08832890	37.4148	8	29
CP x RC Effect	0.32305701	7.5959	8	29

Performance Analyses

System Monitoring Task. Perceptual sensitivity as measured by the non-parametric measure of A' is a common metric for assessing monitoring performance because of its ability to account for a range of user performance, such as number of false alarms and hits. It was used, instead of the variable of a participant's probability of detection used by Singh et al. (1997) because of its sensitivity as a performance measure. There were significant main effects found for both automation reliability condition, F (1,39) = 25.26, p < .0001, and complacency potential, F (1,39) = 16.71, p < .001 (see Table 2). Participants who performed the monitoring task under the variable reliability condition (M = .84) did significantly better than participants under the constant reliability condition (M = .70). This confirms the finding of Parasuraman et al. (1993) that constant reliability, even under high levels of reliability, significantly impairs the ability of the operator to monitor for infrequent automation failures. In addition, high complacency potential participants did significantly worse overall (M = .72) than low complacency potential participants (M = .84). A significant interaction of CP x RC for A', was also found, F (1,39) = 11.49, p < .001. A simple-effects analysis (p < .05) demonstrated that participants across all groups and conditions performed comparably with the exception of the participants in the high complacency potential x constant reliability condition. Figure 2 presents the interaction for CP x RC for A'.

Table 2

Analysis of Variance for Perceptual Sensitivity A’

Source	df	SS	MS	F
Reliability Condition	1	0.21132	0.21132	25.26*
Complacency Potential	1	0.13978	0.13978	16.71*
Complacency Potential X Reliability Condition	1	0.096102	0.096102	11.49*

Note. *p < .001

Tracking Root-Mean-Squared-Error (RMSE) The results of an ANOVA on tracking performance, revealed that participants in the variable reliability condition (M = 28.94) performed significantly worse overall on the tracking task than participants in the constant reliability condition (M = 17.20), F (1, 39) = 28.12, p < .0001. Furthermore, participants assigned to the high complacency potential group (M = 30.15) also had higher tracking RMSE overall than participants in the low complacency potential group (M = 15.98), F (1,39) = 40.89, p < .0001, see Table 3. There was also a significant interaction between Complacency Potential and Reliability Condition, F (1,39) = 8.63, p < .005, for tracking as presented in Figure 3. A simple effects analysis (p < .05) demonstrated that participants classified as high complacency potential in the variable-reliability condition had significantly higher tracking error, than participants in any of the other complacency-reliability combinations. Note that lower RMSE values reflect better tracking performance

Table 3

Analysis of Variance for Tracking Task

Source	df	SS	MS	F
Reliability Condition	1	1380.0816	1380.0816	28.12*
Complacency Potential	1	2007.1983	2007.1983	40.89*
Complacency Potential X Reliability Condition	1	423.61385	423.61385	8.63*

Note. *p < .05

Resource Management . A global measure of task performance was obtained for each participant by computing the RMSE in fuel levels of tanks A and B (deviation from the required level of 2500 gallons). Fuel levels were computed for every 10-min. block to yield the amount of deviation from the required level for each block. An ANOVA, presented in Table 4, did not find any main or interaction effects for resource management performance, p > .05. Participants, across all groups and conditions, did not deviate more than an average of 212 gallons above the criteria of 2500 gallons. This deviation is well within the acceptable range of performance on the resource management task according to Comstock and Arnegard (1992).

Table 4

Analysis of Variance for Resource Management Task

Source	df	SS	MS	F
Reliability Condition	1	11.9629	11.9629	0.00
Complacency Potential	1	3299.697	13299.697	2.85
Complacency Potential X Reliability Condition	1	625.0879	628.0879	0.13

Note. *p < .05

Rating Scales

An ANOVA was conducted for both the NASA-TLX and the Task-related Boredom Scale, which was collected after each experimental session.

Task-Related Boredom Scale (TBS) An ANOVA procedure found a significant main effect for TBS for complacency potential, F (1,39) = 67.31, $p < .0001$. Participants assigned to the high complacency potential group ($M = 25.87$) scored higher on the task-related boredom scale than participants assigned to the low complacency potential group ($M = 14.85$). A significant interaction for complacency potential X reliability condition was also found for the TBS, F (1,39) = 4.58, $p < .05$ (See Figure 4). A simple-effects analysis ($p < .05$) showed that participants high in complacency potential in the constant-reliability condition rated task-related boredom higher than participants low in complacency potential in the constant and variable reliability conditions. The other main effect for reliability was not found to be significant (See Table 5).

Table 5.

Analysis of Variance for Task-Related Boredom Scale

Source	df	SS	MS	F
Reliability Condition	1	21.756250	21.756250	1.20
Complacency Potential	1	1215.5063	1215.5063	67.31*
Complacency Potential X Reliability Condition	1	82.65625	82.65625	4.58*

Note. *p < .05

NASA Task Load Index (NASA-TLX). A significant main effect was found for reliability condition, F (1,39) = 6.82, $p < .01$. Participants in the variable reliability condition rated overall mental workload on the NASA-TLX ($M = 57.05$) to be significantly higher than participants in the constant reliability condition ($M = 46.67$). A main effect was not found for complacency potential, F(1,39) = 3.39, $p > .05$, as seen in Table 6. However, the ANOVA did find a significant interaction for Complacency Potential x Reliability Condition for NASA-TLX, F (1, 39) = 39.93, $p < .0001$. A simple-effects analysis ($p < .05$) showed that participants high in complacency potential rated workload higher in the variable-reliability condition than under any other group-reliability combinations. The interaction is presented in Figure 5.

Table 6

Analysis of Variance for NASA-TLX

Source	df	SS	MS	F
Reliability Condition	1	2175.6250	2175.6250	6.82*
Complacency Potential	1	1081.6000	1081.6000	3.39
Complacency Potential X Reliability Condition	1	12744.900	12744.900	39.93*

Note. *p < .01

Individual Difference Measures

A correlation analysis was conducted to examine the relationship between the three individual difference measures: Cognitive Failure Questionnaire (CFQ), Complacency Potential Rating Scale (CPRS), and the Boredom Proneness Scale (BPS). All three measures were significantly and positively correlated with one another, as seen in Table 7. Correlation analyses were conducted separately for the individual difference measures, total workload scores, total task-related boredom scores, performance on the system- monitoring task for each hypothesis involved.

Table 7
Correlation Analysis of Individual Difference Measures

	BPS	CPRS	CFQ
BPS	1.00000	0.59608	0.71770
	0.0	0.0001	0.0001
CPRS	0.59608	1.00000	0.70945
	.0001	0.0	0.0001
CFQ	0.71770	0.70945	1.00000
	0.0001	0.0001	0.0

Scores on both the Cognitive Failure Questionnaire and the Boredom Proneness Scale were not significantly correlated with performance on the system-monitoring task ($p > .05$). Scores on the CFQ and the BPS were also not significantly correlated with subjective workload as measured by the NASA-TLX ($p > .05$). These correlational analyses are shown in Table 8.

Table 9 shows the correlations between the BPS, CPRS and the Task-related Boredom Scale (TBS). As expected, both BPS and CPRS scores were significantly and positively correlated with TBS score.

Table 8

Correlations Among Total Reported Workload, and Perceptual Sensitivity (A') and Individual Difference Measures

	TLXTOTAL	A'
CFQ	0.02859 0.8610	-0.25662 0.1099
BPS	0.23084 0.1518	-0.25600 0.1108

Table 9

Correlations Among Total Task-Related Boredom and Individual Difference Measures

	BPS	CPRS
TBSTOTAL	0.63679 0.0001	0.52099 0.0006

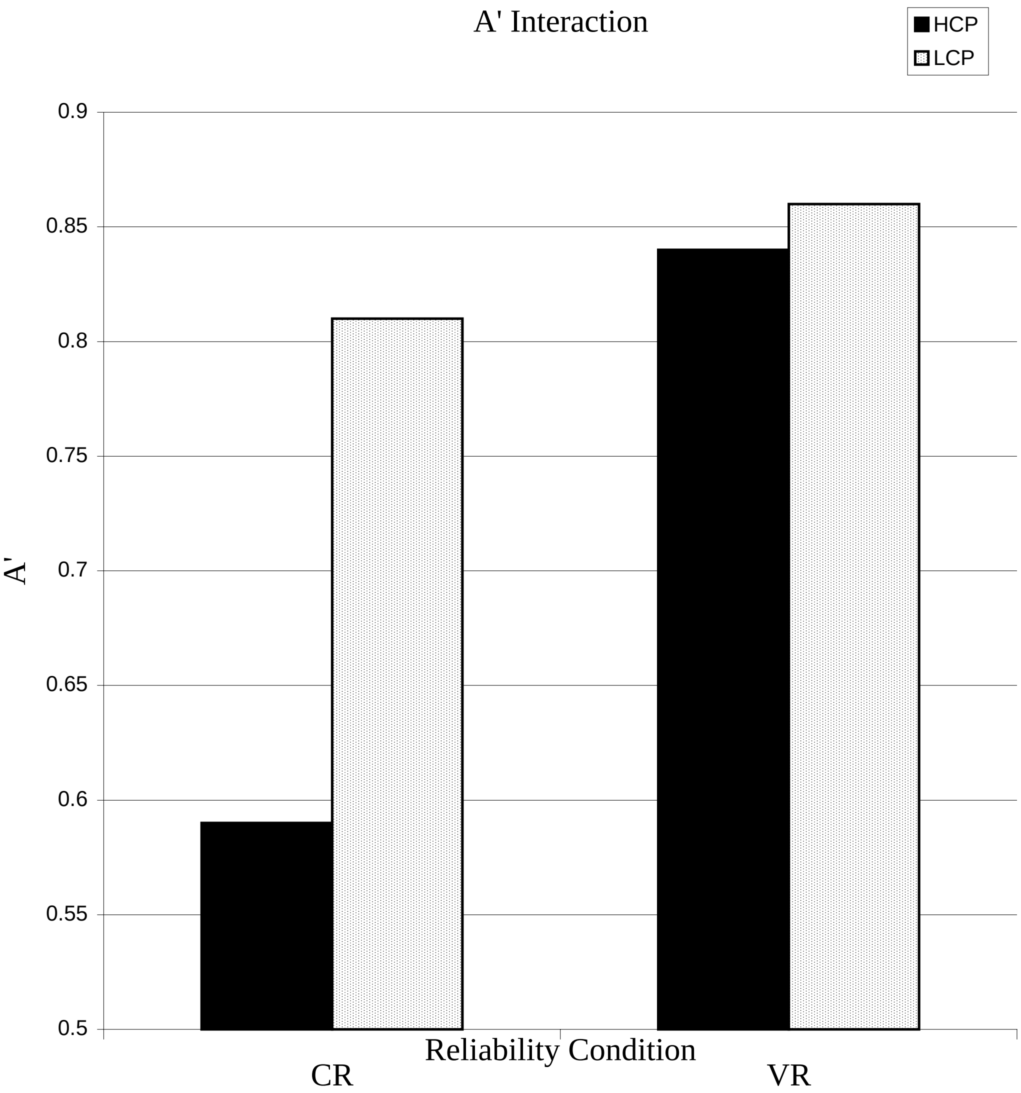

A' Interaction
■ HCP
▨ LCP
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
A'
Reliability Condition
CR
VR

Figure 2.

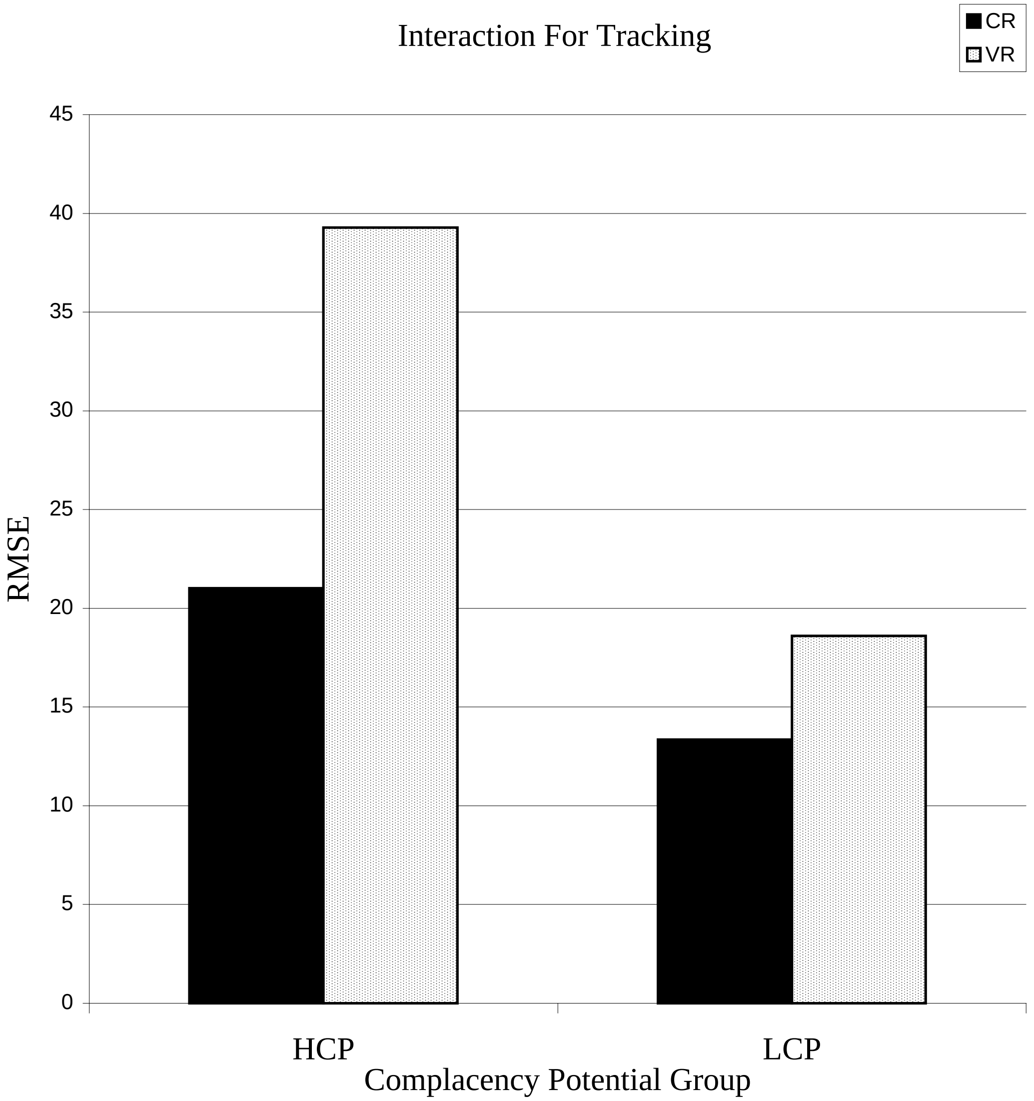

Interaction For Tracking
■ CR
▨ VR
45
40
35
30
25
20
15
10
5
0
RMSE
HCP
LCP
Complacency Potential Group

Figure 3.

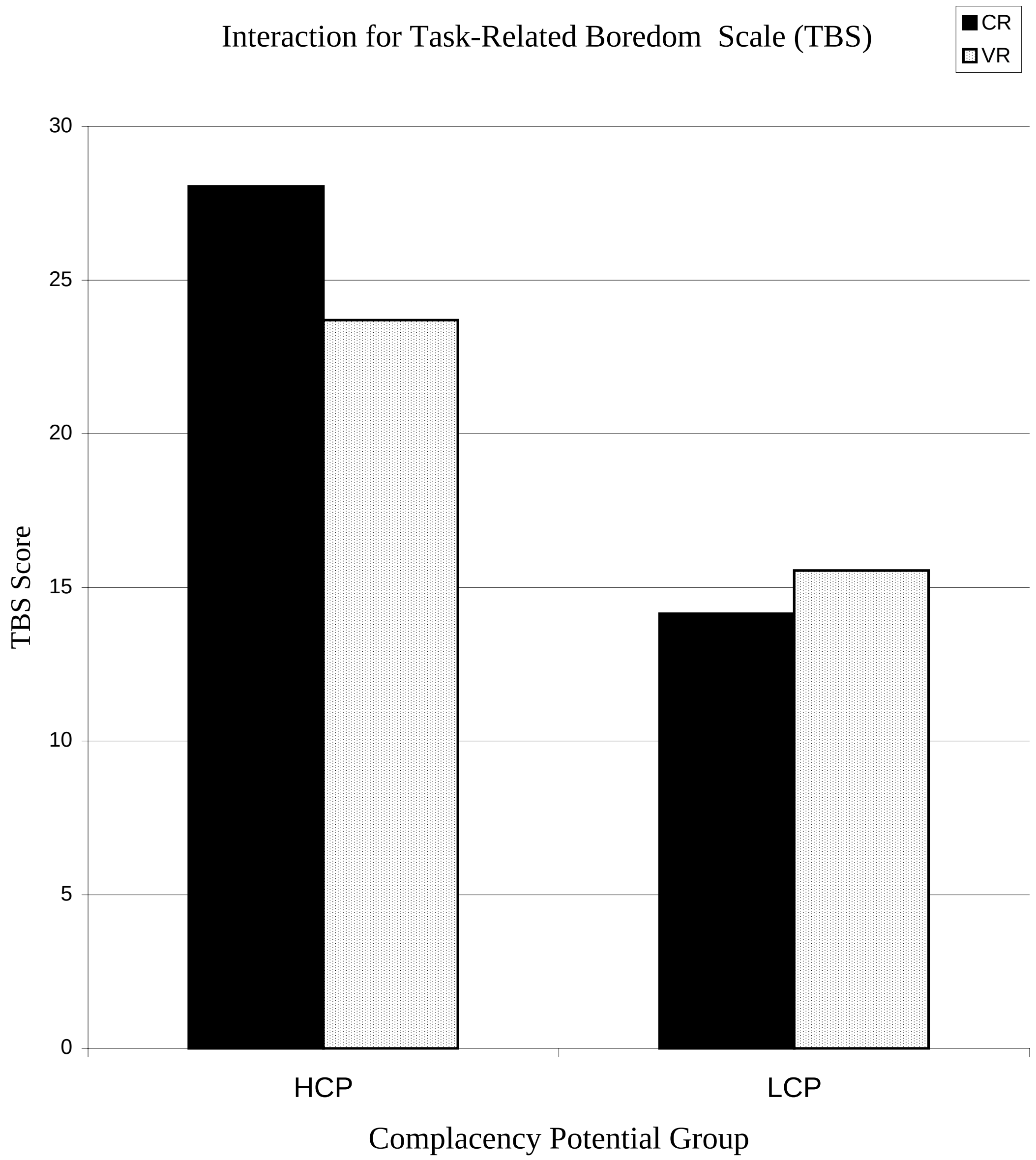

Interaction for Task-Related Boredom Scale (TBS)
■ CR
▨ VR
30
25
20
15
10
5
0
TBS Score
HCP
LCP
Complacency Potential Group

Figure 4.

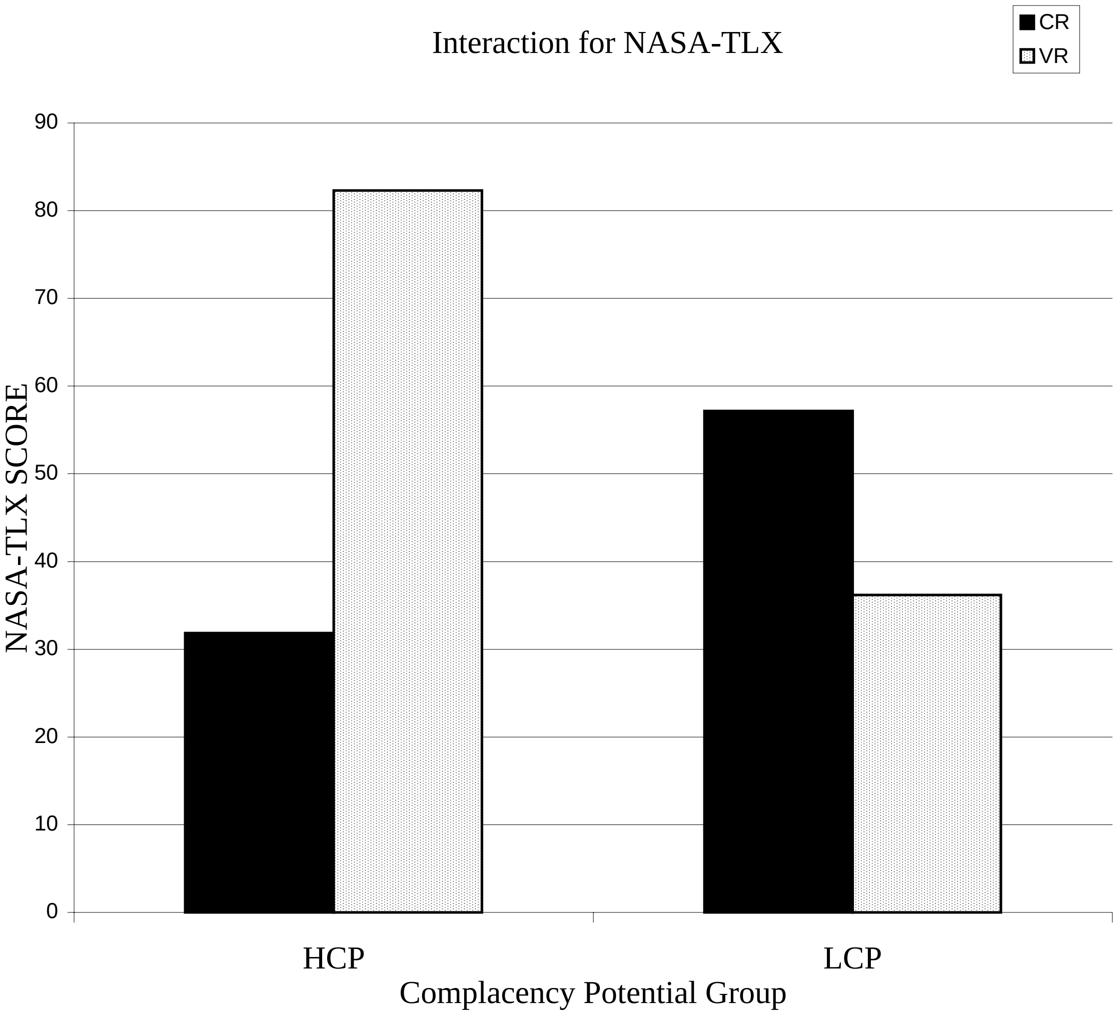

Interaction for NASA-TLX
■ CR
▨ VR
NASA-TLX SCORE
90
80
70
60
50
40
30
20
10
0
HCP
LCP
Complacency Potential Group

Figure 5.

DISCUSSION

An increasing number of modern work environments are at least partially, if not extremely, dependent on some form of automation. Nowhere is this more prevalent than the field of modern aviation. Automation has been implemented in all modern airplanes in order to make air travel safer and more efficient. Its widespread use and advancement in the cockpit has replaced pilots as the main operator of airplanes and has placed them in the role of system monitors. Pilots, however, are often unable to effectively monitor automated tasks and may enter into a “hazardous state of awareness”, which has been documented to be a major contributor of aviation accidents (FAA, 1996). One of the most prevalent “hazardous states of awareness” involved in such accidents is automation-induced complacency.

Complacency is believed to occur when an operator must monitor an automated system and detect possible failures within a multi-task environment (Parasuraman et al., 1993). Although researchers agree that it is a major problem with aviation, little consensus is held as to exactly what complacency is, what the best methods of measuring it are, and how to combat it in the modern cockpit. Riley (1996) noted that there are many possible individual and social factors involved in complacency experienced by pilots, including overreliance, trust, workload, and boredom that are not understood in relation to the effects these psychological factors have on complacency. Therefore, before the aviation industry can begin to implement remediations to what is widely known to be a significant obstacle for automation’s potential for increasing aviation safety, it is necessary to fundamentally understand the individual difference variables of the operator in order to have a complete picture of this “hazardous state of awareness”.

The present study was designed to begin to accomplish these research goals, and to examine the impact that previous research has postulated may be an underlying component of automation-induced complacency. A between-subjects design was used in which group-selected participants, based on their scores on the Complacency Potential Scale (Singh, Parasuraman, & Molloy, 1993a) were randomly assigned to two automation reliability conditions. 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. Participants in the constant-reliability group experienced the automation reliability as a constant schedule from block to block at 87.5% (14 out of 16 malfunctions detected by the automation) for each of the participants. Half of the participants in the variable-reliability group experienced the automation reliability alternating every 10 min. block from low (56.25%, 9 out of 16 malfunctions detected) to high (87.5%), and the other half of the participants in this group experienced the reliability of the automation alternating from high to low. These reliability schedules have been used in other studies (e.g., Parasuraman, Molloy, & Singh, 1993; Singh et al., 1997) and have been shown to produce the necessary psychological conditions to induce complacency.

Task Performance

Secondary Task Performance. Parasuraman, Molloy, and Singh (1993) defined complacency as “…the operator failing to detect a failure in the automated control of a system-monitoring task" (p. 4). They reasoned that complacent behaviors arise during high workload conditions and that complacency is an outcome of task / attention allocation strategies. However, other studies (Thackary & Touchstone, 1989; Lee & Moray, 1992; Muir & Moray, 1996) have posited that it is a combination of automation and operator intra-personal characteristics that determine whether an operator may adopt a particular strategy. Parasuraman, Molloy, & Singh (1993) reported that automation-induced complacency cannot be produced under single-condition, low workload environments. Wiener (1981) further illustrated, through over 500 incidents in the ASRS, that complacency could be attributed to an overreliance on automated systems. These complacent behaviors become evident during high workload conditions in which the automation fails "strong and silent" and the pilot is "…in nonvigilance based on an unjustified assumption of satisfactory system-state." Therefore, automation-induced complacency behaviors occur under highly reliable and high workload conditions, in which the operator's confidence in the automation produces an attention allocation strategy to "trust" the automation. Based on such an operator definition, we hypothesized that participants in the constant-reliability condition would experience complacency as indexed by poorer performance (A') on the system-monitoring task compared to participants in the variable-reliability condition.

There was a significant main effect found for reliability condition. Participants performing the monitoring task under the variable-reliability condition did significantly better than those participants in the constant-reliability condition. The poor performance of the constant reliability participants confirms the findings of Parasuraman et al. (1993) that a constant reliability schedule, especially under high levels of automation reliability, impairs an operator's ability to monitor for infrequent automation failures within a multitask environment.

All individuals, however, do not trust automation and therefore do not succumb to automation-induced complacency under the conditions just described. As pilots are human beings, there are individual differences, which predispose them to behaving in certain ways in different situations. The Complacency-Potential Rating Scale was developed to measure an individual's predisposition towards becoming complacent when working with automated systems. Singh and his colleagues (1993b) reported that a person who scores high on complacency potential would be more likely to show poor backup monitoring of automation failures under conditions encouraging complacency (i.e., under constant-automation reliability) but not under conditions that did not induce complacency (i.e. variable-reliability).

Similarly, our results supported the findings of Singh et al. (1993b) through a significant interaction of reliability condition by complacency potential for A'. Participants with high complacency potential (HCP) in the constant reliability condition did significantly worse than participants in the other three conditions (i.e. low complacency potential x variable reliability, high complacency potential x variable reliability condition, and low complacency potential x constant reliability condition). High complacency individuals in the constant reliability condition may have trusted the automation and, therefore, these participants missed most of the system failures, leading to a lower A' score. Those who had low complacency potential (LCP) scores, regardless of the reliability condition, did not trust the automation and, therefore, were more vigilant of the automation failures.

Primary Task Performance. The results of the tracking data demonstrated that participants in the constant-reliability condition performed the tracking task significantly better than participants in the variable-reliability condition. Mosier, Skitka, & Korte (1994) stated that most aviation monitoring failures occur when a pilot is engaged in a multi-task situation because they are time-sharing attention allocation across various sub-tasks, the result of which is a shift from multiple- to single-task performance under high workload conditions. This suggests that strategy selection can change workload in a multi-task environment so that efficient strategies, such as trusting the automation, may lower workload. Incidents in the ASRS have numerous accounts in which general aviation pilots focus in on the “T” instruments and “peripheralize” other task demands as a workload management strategy. Therefore, automation-induced complacent behaviors may actually improve performance on other tasks because of the “automation trust” that therein allows the automation to perform that task and frees up cognitive resources to manage other tasks. Hence, performance would be significantly better for the primary task(s) since there was no automated aiding, as was the case with participants in the constant-reliability condition.

Analysis of the resource management task, however, showed no statistical significance. This is not surprising since it is a strategy task that requires few cognitive resources as a primary task (Comstock and Arnegard, 1992). Other studies (Parasuraman, Singh, & Molloy, 1993; Singh, Molloy, & Parasuraman, 1997) have also reported no differences in resource management as a function of automation reliability condition. Unlike the system-monitoring task which requires constant vigilance and system monitoring in order to “catch” system failures, the resource-management task allows for varied response strategies and time-sharing, such as simultaneous responding, alternative responding, or massed responding (Damos, Smist, & Bittner, 1983). Participant post-experimental discussions in the current study voiced that a massed response strategy was often used, allowing the participants to maintain overall successful performance with intermittent periods of performance lapses.

Subjective Rating Scales

Task-Related Boredom. There have been few studies that have discussed the role of boredom, and the present study represents the first known study to empirically examine the construct of task-related boredom and its relationship to automation-induced complacency. Parasuraman, Molloy, & Singh, (1993), however, did state that complacency was distinct from boredom because the single task condition in their study did not produce complacent behaviors; the single-task condition was thought to be a low workload task and, therefore, a boring task. It should be noted that they did not gather any workload or boredom data to substantiate the claim. Furthermore, the ASRS contains numerous descriptions of crews becoming “complacent” because of “boredom” and lowered vigilance and lapsing attention (Pope & Bogart, 1992). Therefore, the earlier claims made by Parasuraman and his colleagues that complacency and boredom are unrelated constructs are not as yet warranted and needed to be examined further.

The current study disagrees with Parasuraman et al.’s (1993) claim that boredom and complacency are distinct concepts. Results of the analysis on the task-related boredom data indicated that HCP participants in the constant-reliability condition rated task-related boredom higher than LCP participants in both constant and variable-reliability conditions. Therefore, the data from this study would suggest that the relationship between complacency and boredom may be more than a casual one. Those participants rated high in complacency potential did in fact experience significantly higher levels of boredom when put in an environment that induces complacency (i.e., constant reliability condition) whereas participants rated low in complacency potential did not experience significantly high levels of task-related boredom. Moreover, the observation was not due solely to whether one was rated high or low in complacency potential. Participants rated high in complacency potential but who performed the tasks in the variable-reliability condition did not report high levels of task-related boredom. These HCP participants also did not show task performance that would suggest that they were experiencing complacency under the variable reliability condition. In other words, only those participants who were predisposed to engage in complacent behaviors (i.e., HCP) and who performed the task under task conditions known to induce complacency (i.e., constant reliability condition) showed signs of complacency (i.e., poor A' scores) and rated task-related boredom significantly high. These participants have the predisposition to trust the automation and, therefore, handed off the system-monitoring task to the automation leaving them with only two, rather than three, tasks to perform. Because the automation was "behaving" and seemed to be performing correctly (i.e., because of the constant reliability schedule), these participants were free to do so. However, the other HCP participants performing the MAT under the variable reliability condition could not hand off the task to the automation because the schedule made it obvious to the participant that automation could not be trusted. Therefore, these participants did not rate the task as high in task-related boredom and did significantly better in terms of A' on the system monitoring task, but reported significantly higher subjective workload.

Subjective Workload. Parasuraman, Molloy, and Singh (1993) noted that automation-induced complacency only arises under conditions of high workload. The point made was that complacency might reflect an "attitude" towards automation that allows them to "trust" the automation as a strategy for dealing with the high workload. Of course, perception of workload and actual "certified" workload is different (Wise & Hopkins, 2000). Workload can be perceived to be different by different pilots, which is why it is called "subjective workload". How each pilot perceives the workload level will determine how he or she responds to the task situation and what strategy they may employ to deal with those cognitive demands.

The present study suggests that perception of workload and automation-induced complacency was determined largely on the basis of whether the participant was classified as high or low in complacency potential. There was a significant main effect found for reliability condition in which participants in the variable reliability condition rated overall workload significantly higher than participants in the constant reliability condition. The finding would run counter to the claims made by other researchers that high workload is a necessary component for automation-induced complacency. In fact, the data suggests that there is no difference in workload rating between variable and constant reliability conditions, despite the different automation schedules. The main effect, instead, is the result of the high workload scores for the high complacency potential subjects.

HCP participants in the variable reliability condition rated workload significantly higher than the LCP participants, in both the variable and constant reliability conditions, and HCP participants in the constant reliability condition. Like the boredom results, these results suggest that individual differences between HCP and LCP participants were the significant factors which operated to determine the onset of automation-induced complacency. This may be due to the strategy different individuals employed in order to cope with the workload of performing the tasks on the MAT battery. The HCP participants, who performed the tasks under constant-reliability condition, may have trusted the automation in the system-monitoring task and therefore may have only been practically performing two tasks; this would also account for why they reported task-related boredom higher. Those low in complacency potential did not ever trust the automation and therefore, relatively speaking didn't statistically report a difference in workload between the two reliability conditions. HCP participants, on the other hand, have a predisposition toward trusting the automation and it requires a great deal of "cognitive overhead" to decide not to trust and monitor the automation. The HCP and variable reliability participants, having noticed that the automation was not perfect and acted erratically, had to then monitor the automation because it was obvious that the automation could not be trusted. This ran counter to their individual difference strategy predisposition to trust the automation and, therefore, participants in this group reported significantly higher workload than the other complacency potential, reliability condition groups.

Correlational Analysis of Individual Difference Variate Measures

Results from the correlational analysis demonstrated that participants' complacency potential was strongly related to their level of cognitive failure and proneness to boredom. As discussed above, the Complacency Potential Rating Scale (CPRS) measures an individual's propensity to exhibit complacent behaviors. The scale measures attitudes towards everyday automated devices, such as automobile cruise controls and automatic teller machines. The scale has been shown to index a person's trust in, reliance on, and confidence in automation, and that these are the major determinants of automation-induced complacency propensity.

Singh and his colleagues (1993b) view automation-induced complacency as an attitude toward automation rather than as a state or trait. They concluded that there was no evidence to support the idea that complacency is a psychological state that is experienced by pilots. They also concluded that there is no clear indication as to whether complacency is an enduring trait experienced by some individuals. Pilot reports, however, tend to disagree with Singh and his colleagues and instead support the view that complacency is a psychological state or an intervening variable that is influenced by psychological states, such as boredom or fatigue. Flight crews often complain of becoming "complacent" because of "boredom" and as a result an operator's proneness to boredom was examined in the current study (Pope & Bogart, 1992). The personality dimension "cognitive failure" may also be a precursor to an individual's becoming complacent while monitoring automation. According to Grubb et al. (1994), operators who are high in cognitive failure tend to be more absent-minded, forgetful, error prone, and less able to allocate mental resources to perform monitoring tasks than those classified as low cognitive failure. The present study wanted to more closely examine the relationship between complacency potential, boredom proneness and cognitive failure, to better define what complacency is.

The strong relationship between the three individual difference variables shows that the Boredom Proneness Scale (BPS) and the Cognitive Failure Questionnaire (CFQ) tap into some of the same properties as the Complacency Potential Rating Scale. Thus supporting the overarching view of the current study that cognitive failure and boredom proneness are part of the complex psychological phenomenon of complacency.

Scores on the CFQ and BPS, however, did not demonstrate the expected significant relationship with performance on the system-monitoring task or with subjective workload as measured by the NASA-TLX (see Table 2). Therefore, these particular scales may not be good predictors of performance on the system-monitoring task. Nevertheless, the two scales were highly correlated with the CPRS and there was a significant interaction of Complacency Potential x Reliability Condition for A' scores on the system monitoring task. The results of this interaction demonstrated that participants across all groups and conditions performed comparably with the exception of the participants in the high complacency x constant reliability condition who performed significantly worse (see Figure 5). Scores on the BPS and the CPRS did demonstrate a strong relationship with the task-related boredom scale. Thus, as expected those high in boredom proneness (and high in complacency potential) did experience significantly higher reported boredom. There was also a significant interaction for CP X RC for the task-related boredom scale, which demonstrated that participants high in complacency potential in the constant-reliability (who became complacent) rated task-related boredom higher than participants in any other group-reliability condition did. The significant correlations to CPRS and the interaction effects of CP x RC for the TBS and A', continue to support the idea that cognitive failure and boredom proneness may be a part of the psychological construct of complacency. The lack of significant relationship found by the correlational analyses on these two measures with system monitoring performance does not imply that they are unrelated to automation-induced complacency. For the present study's findings may be a result of the study design (e.g. correlation analysis) not being sensitive enough to tease out these elements when examining them in relation to actual task performance.

Weaknesses of the Present Study

Although there are a number of important strengths of the present research, several limitations should also be noted. One such weakness is its use of automation that utilizes an extremely high degree of unreliability (i.e. 87.5% reliable) that would never be tolerated in a real world environment. Real world automation has a reliability of 99.99%; in the case of aviation there are several back-up systems to catch system-failures before they ever become apparent to the human operator. However, even with this unrealistic degree of reliability there were still strong significant effects shown by the present research, thus demonstrating that the problem and effects of automation-induced complacency are even more problematic in real world settings. Real world aviation is much more complicated than the MAT-battery and requires trained pilots to operate it, rather than novice undergraduate students as used in the current study. The extensive training of pilots, however, does not safeguard them from experiencing complacency and sometimes can be one of the major factors leading to its occurrence as the pilots have come to trust the automation they have been trained to use. The present study was also unable to tease out the individual difference variables of cognitive failure and boredom proneness, due to its design, in order to find how they contribute to the onset of complacency. Nevertheless, measures for both variables were strongly correlated with the Complacency Potential Rating Scale, indicating that they are highly related to the complex psychological construct of complacency and therefore should be considered in future research efforts.

Future Research

This study has shown that there are personality individual differences that are related to whether an individual will succumb to automation-induced complacency. However, it was unable to tease out how significant the effects of such personality variables as boredom proneness and cognitive failure are on this psychological state. Future research should classify the entire subject pool on each of these variables and then randomly assign them to groups, in order to conclude exactly what type of an effect these personality variables have on the occurrence of complacency. Other individual difference variables may also need to be considered. Perhaps additional research could be done with actual pilots in order to examine how their personality differences affect them when using automation, and if certain variables make them more prone to experiencing this state. Thus, due to the exploratory nature of the present study, future research should focus on expanding all points that have been examined.

The conclusion that may be drawn from this study is that complacency is a psychological state that is induced by personality predispositions that influence human-automation interaction. Although these various scales have significant cross-correlations and perhaps measure many of the same underlying personality psychometrics, the results add to the growing body of evidence that complacency is a highly complex psychological construct within the field of aviation that warrants further study.

REFERENCES

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.

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 Earlbaum Associates.

APPENDIX A

Please read each statement carefully and circle the one response that you feel most accurately describes your views and experiences. THERE ARE NO RIGHT OR WRONG ANSWERS. Please answer honestly and do not skip any questions.

SA	A	U	D	SD
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

SA A U D SD 1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.

SA A U D SD 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.

SA A U D SD 3. People save time by using automatic teller machines (ATMs) rather than a bank teller in making transactions.

SA A U D SD 4. I do not trust automated devices such as ATMs and computerized airline reservation systems.

SA A U D SD 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.

SA A U D SD 6. I feel safer depositing my money at an ATM than with a human teller.

SA A U D SD 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.

SA A U D SD 8. People whose jobs require them to work with automated systems are lonelier than people who do not work with such devices.

SA A U D SD 9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journey safer.

SA A U D SD 10. ATMs provide safeguard against the inappropriate use of an individual’s bank account by dishonest people.

SA A U D SD 11. Automated devices used in aviation and banking have made work easier for both employees and customers.

SA A U D SD 12. I often use automated devices.

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

SA A U D SD 14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.

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

SA A U D SD 16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.

SA A U D SD 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.

SA A U D SD 18. Work has become more difficult with the increase of automation in aviation and banking.

SA A U D SD 19. I do not like to use ATMs because I feel that they are sometimes unreliable.

SA A U D SD 20. I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.

APPENDIX B

Instructions: Circle the answer that is most consistent with your attitudes or beliefs about yourself. You MUST select either “yes” or “no”. THERE ARE NO RIGHT OR WRONG ANSWERS. Please do not skip any questions.

1. It is easy for me to concentrate on my activities yes / no
2. Frequently when I am working I find myself worrying about other things yes / no
3. Time always seems to be passing slowly yes / no
4. I often find myself at “loose ends” not knowing what to do yes / no
5. I am often trapped in situations where I have to do meaningless things yes / no
6. Having to look at someone’s home movies bores me tremendously yes / no
7. I have projects in mind all the time, things to do yes / no
8. I find it easy to entertain myself yes / no
9. Many things I have to do are repetitive and monotonous yes / no
10. It takes more stimulation to get me going than most people yes / no
11. I get a kick out of most things I do yes / no
12. I am seldom excited about my work yes / no
13. In any situation I can usually find something to do or see to keep me interested yes / no
14. Much of the time I just sit around doing nothing yes / no
15. I am good at waiting patiently yes / no
16. I often find myself with nothing to do, time on my hands yes / no
17. In situations where I have to wait, such as in line, I get very restless yes / no
18. I often wake up with a new idea yes / no
19. It would be very hard for me to find a job that is exciting enough yes / no
20. I would like more challenging things to do in life yes / no
21. I feel that I am working below my abilities most of the time yes / no
22. Many people would say that I am a creative or imaginative person yes / no
23. I have so many interests, I don’t have time to do everything yes / no
24. Among my friends, I am the one who keeps doing something the longest yes / no
25. Unless I am doing something exciting, even dangerous, I feel half-dead and dull yes / no
26. It takes a lot of change and variety to keep me really happy yes / no
27. It seems that the same things are on television or the movies all the time; it’s getting old yes / no
28. When I was young, I was often in monotonous and tiresome situations yes / no

APPENDIX C

	Very Often	Quite Often	Very Occasionally	Rarely	Never
1.Do you read something and find you haven't	4	3	2	1	0
been thinking about it and must read it again?	4	3	2	1	0
2. Do you find you forget why you went from					
one part of the house to the other?	4	3	2	1	0
3. Do you fail to notice signposts on the road?	4	3	2	1	0
4. Do you find you confuse right and left when					
giving directions?	4	3	2	1	0
5. Do you bump into people?					
6. Do you find you forget whether you've turned	4	3	2	1	0
off a light or a fire or locked the door'?					
7. Do you fail to listen to people's names when	4	3	2	1	0
you are meeting them?					
8. Do you say something and realize afterwards	4	3	2	1	0
that it might be taken as insulting?	4	3	2	1	0
9. Do you fail to hear people speaking to you					
when you are doing something else?	4	3	2	1	0
10. Do you lose your temper and regret it?					
11. Do you leave important letters unanswered	4	3	2	1	0
for days?					
12. Do you find you forget which way to turn on	4	3	2	1	0
a road you know well but rarely use?					
13. Do you fail to see what you want in a	4	3	2	1	0
supermarket (although it's there)?	4	3	2	1	0
14. Do you find yourself suddenly wondering	4	3	2	1	0
whether you've used a word correctly?					
15. Do you have trouble making tip your mind?	4	3	2	1	0
16. Do you find you forget appointments?					
17. Do you forget where you put something like					
newspaper or a book?					
18. Do you find you accidentally throw away the					
thing you want and keep what you meant to	4	3	2	1	0
throw away - as in the example of throwing away					
the matchbox and putting the used match in your	4	3	2	1	0
pocket?	4	3	2	1	0
19. Do you daydream when you ought to be					
listening to something?					
20. Do you find you forget people's names'?	4	3	2	1	0
21. Do you start doing one thing at home and get					
distracted into doing something else					
unintentionally?	4	3	2	1	0
22. Do you find you can't quite remember					
something although it's 'on the tip of your	4	3	2	1	0
tongue?	4	3	2	1	0
23. Do you find you forget what you came to the					
shops to buy?	4	3	2	1	0
24. Do you drop things?	4	3	2	1	0
25. Do you find you can't think of anything to					
say?					

APPENDIX D

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

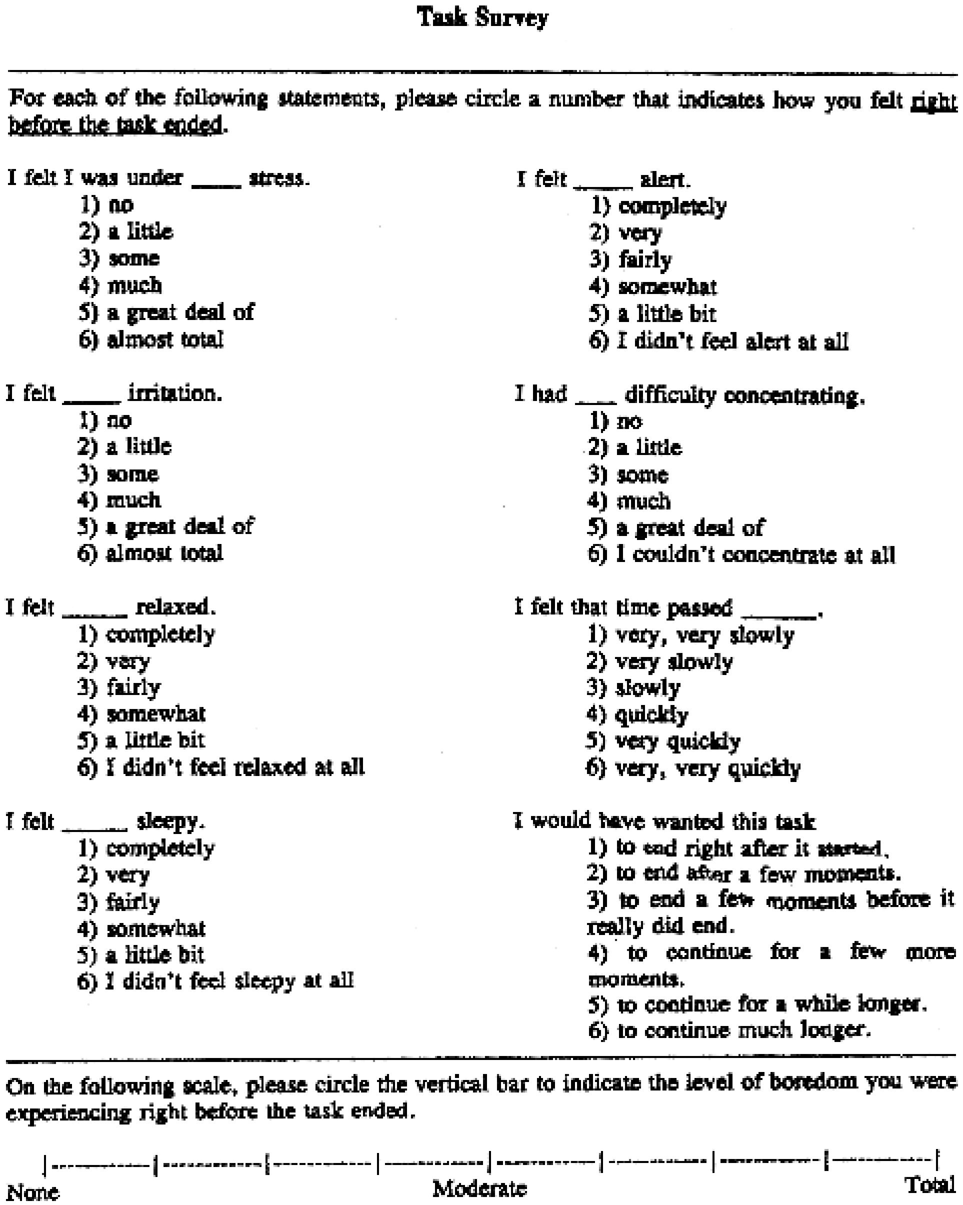
EFFORT

Low High

FRUSTRATION

Low High

APPENDIX E



Task Survey

For each of the following statements, please circle a number that indicates how you felt <u>right before the task ended</u>.

I felt I was under ____ stress.
1) no
2) a little
3) some
4) much
5) a great deal of
6) almost total

I felt _____ irritation.
1) no
2) a little
3) some
4) much
5) a great deal of
6) almost total

I felt ______ relaxed.
1) completely
2) very
3) fairly
4) somewhat
5) a little bit
6) I didn't feel relaxed at all

I felt _____ sleepy.
1) completely
2) very
3) fairly
4) somewhat
5) a little bit
6) I didn't feel sleepy at all

I felt _____ alert.
1) completely
2) very
3) fairly
4) somewhat
5) a little bit
6) I didn't feel alert at all

I had ___ difficulty concentrating.
1) no
2) a little
3) some
4) much
5) a great deal of
6) I couldn't concentrate at all

I felt that time passed ______.
1) very, very slowly
2) very slowly
3) slowly
4) quickly
5) very quickly
6) very, very quickly

I would have wanted this task
1) to end right after it started.
2) to end after a few moments.
3) to end a few moments before it really did end.
4) to continue for a few more moments.
5) to continue for a while longer.
6) to continue much longer.

On the following scale, please circle the vertical bar to indicate the level of boredom you were experiencing right before the task ended.

|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

None Moderate Total

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
December 2001

3. REPORT TYPE AND DATES COVERED
Technical Memorandum

4. TITLE AND SUBTITLE
Examination of Automation-Induced Complacency and Individual Difference Variates

5. FUNDING NUMBERS
WU 711-50-21-01

6. AUTHOR(S)
Lawrence J. Prinzel III, Holly DeVries, Fred G. Freeman, and Peter Mikulka

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L-18140

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-2001-211413

11. SUPPLEMENTARY NOTES
Prinzel: Langley Research Center, Hampton, VA; DeVries, Freeman, and Mikulka, Old Dominion University, Norfolk, VA

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 54 Distribution: Nonstandard
Availability: NASA CASI (301) 621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Automation-induced complacency has been documented as a cause or contributing factor in many airplane accidents throughout the last two decades. It is surmised that the condition results when a crew is working in highly reliable automated environments in which they serve as supervisory controllers monitoring system states for occasional automation failures. Although many reports have discussed the dangers of complacency, little empirical research has been produced to substantiate its harmful effects on performance as well as what factors produce complacency. There have been some suggestions, however, that individual characteristics could serve as possible predictors of performance in automated systems. The present study examined relationship between the individual differences of complacency potential, boredom proneness, and cognitive failure, automation-induced complacency. Workload and boredom scores were also collected and analyzed in relation to the three individual differences. The results of the study demonstrated that there are personality individual differences that are related to whether an individual will succumb to automation-induced complacency. Theoretical and practical implications are discussed.

14. SUBJECT TERMS
Complacency, Automation, Personality, Individual Differences, Overreliance

15. NUMBER OF PAGES
48

16. PRICE CODE
A03

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z-39-18
298-102