THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 3(1), 1–23

FORMAL PAPERS

Performance Consequences of Automation-Induced "Complacency"

Raja Parasuraman, Robert Molloy, and Indramani L. Singh
Cognitive Science Laboratory
The Catholic University of America
Washington, DC

The effect of variations in the reliability of an automated monitoring system on human operator detection of automation failures was examined in two experiments. For four 30-min sessions, 40 subjects performed an IBM PC–based flight simulation that included manual tracking and fuel-management tasks, as well as a system-monitoring task that was under automation control. Automation reliability—the percentage of system malfunctions detected by the automation routine—either remained constant at a low or high level over time or alternated every 10 min from low to high. Operator detection of automation failures was substantially worse for constant-reliability than for variable-reliability automation after about 20 min under automation control, indicating that the former condition induced "complacency." When system monitoring was the only task, detection was very efficient and was unaffected by variations in automation reliability. The results provide the first empirical evidence of the performance consequences of automation-induced "complacency." We relate findings to operator attitudes toward automation and discuss implications for cockpit automation design.

Cockpit automation has brought increased demands on crews to monitor automated systems for malfunctions or failures (Chambers & Nagel, 1985; D. A. Norman, 1991; R. Parasuraman, 1987; Wiener, 1987, 1988; Wiener & Curry, 1980). Al-

Requests for reprints should be sent to Raja Parasuraman, Cognitive Science Laboratory, The Catholic University of America, Washington, DC 20064.

though some monitoring functions can themselves be automated (see, e.g., Palmer & Degani, 1991), crew monitoring will be required as long as aircraft cockpits are manned and machine monitoring of aircraft systems is less than perfect. "Backup" monitoring by controllers will also be required in fully automated air traffic control (ATC) concepts such as Automated En Route ATC, which may be implemented under the National Airspace Plan (Kingsbury, 1986; R. Parasuraman, 1987; Thackray & Touchstone, 1989).

Crew "complacency" is often mentioned as representing one potential negative effect of automation relevant to monitoring performance (Singh, Molloy, & R. Parasuraman, in press; Thackray & Touchstone, 1989; Wiener, 1981). The condition is thought to arise in highly reliable automated environments in which the pilot or the controller serves in a backup role. Currently, however, there is little consensus regarding the definition of *complacency.* Wiener (1981) defined it as "a psychological state characterized by a low index of suspicion" (p. 117), and Billings, Lauber, Funkhouser, Lyman, and Huff (1976), in the NASA 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" (p. 23). Despite the lack of consensus on usage and terminology, pilot complacency has long been implicated as a possible contributing factor in aviation accidents (Hurst & Hurst, 1982; Wiener, 1981). The term has become more prominent with the advent of automated aircraft and with the attendant possibility of automation-induced complacency.

Although researchers agree that complacency may be a potential problem in automated aircraft, there is considerable uncertainty regarding the dimensions of complacency (but cf. Singh et al., in press, which is discussed later). Nevertheless, the frequency with which the term *complacency* is encountered in analyses of aviation accidents suggests that attempts should be made to understand and validate the concept. Wiener (1981) proposed that empirical research and analyses of data on aviation incidents are necessary to gain an understanding of the mechanisms of complacency.

Singh et al. (in press) recently developed the 20-item Complacency Potential Rating Scale in assessing attitudes toward everyday automated devices such as automated teller machines, automobile cruise controls, and so on. The internal consistency and test–retest reliability of the scale were found to be high. Factor analysis of sample responses ($N = 139$) indicated that the major factors contributing to "complacency potential" were a person's trust in, reliance on, and confidence in automation. Singh et al. suggested that high scores on these factors could be associated with complacency. The ASRS database includes several examples of aviation incidents attributed to crew overconfidence in and overreliance on automated systems. The following is one example:

> While flying, the first officer (F/O) programmed the Flight Management Computer (FMC) to guide the autopilot to cross DRAKO intersection below FL230 and above FL170 at 250 knots. The right autopilot was in use in the command position and the mode control panel (MCP) was set in LNAV and VNAV. After descending to FL260

> ATC cleared the flight for the profile descent into Denver. The F/O put 17,000 feet in the altitude box on the MCP. At approximately FL220 the pilot tuned in Denver ATIS. The radios started to get scratchy as the flight entered a band of turbulence. The crew's radar did not indicate a heavy cell, only light precipitation. The captain turned on the engine and wing anti-ice systems and shortly thereafter an "ice detect" light indicated on the crew's EICAS CRT. Approaching 19,000 feet, the aircraft received heavy static discharges and more turbulence. Eight miles from the DRAKO intersection ATC informed the crew to climb immediately to avoid high terrain. The captain immediately took over flying the aircraft, applied power and climbed from 15,000 feet to FL170. The crew attributed this altitude deviation to high workload and to over-reliance on the FMC. (Billings et al., 1976)

Crew attitudes such as overconfidence in automation may not be sufficient in themselves to lead to complacency but may only indicate a potential for complacency. Complacent behavior may arise only when complacency potential occurs jointly with other conditions such as high workload brought about by poor weather, heavy traffic, or fatigue due to poor sleep or long flights. The combination of the crew's attitude toward automation (e.g., overreliance) and a particular situation (e.g., fatigue) may lead to complacent behavior. One index of complacent behavior (among other possibilities) could be reduced accuracy or delay in detecting a failure in the automated control of a flight task. However, there is currently little empirical evidence concerning the performance consequences of automation-related complacency.

This article describes what we believe to be the first such empirical evidence. In Thackray and Touchstone (1989), subjects performed a simulated ATC task either with or without the help of an automated aid that provided advisory messages concerning potential aircraft-to-aircraft conflicts. The automation failed twice, early and late during a 2-hr session. Thackray and Touchstone (1989) reasoned that subjects using the automated aid may become complacent and thus fail to detect the failures. Although subjects were somewhat slower to respond to the first failure when using the automated aid, this was not the case for the later failure. In general, subjects were as efficient at monitoring in the presence of automation as they were in its absence.

Thackray and Touchstone (1989) indicated that their failure to obtain reliable evidence of complacency might be related to their use of a relatively short test session (2 hr). Another factor might be that subjects in this study were responsible for only a single task, that of monitoring. We reasoned that the performance consequences of automation-induced complacency might be more likely when the operator is responsible for many functions, not just for the monitoring function (cf. Billings et al.'s, 1976, ASRS report of altitude deviation in which the crew was involved in many tasks when the automation failed). In such an environment, characteristics of the automated device carrying out a function (e.g., its reliability and consistency) might influence how well the operator is able to detect and respond to an automation failure. We therefore examined the efficiency of human monitoring of system failures in a multitask environment. Using an IBM PC–based flight-simulation task, we had subjects perform manual tracking and fuel-management tasks in addition to a system-monitoring task, which was carried out by automation that was not perfectly reliable. Complacency was defined as the operator failing to detect a failure in the automated control of the system-monitoring task.

Langer (1989) introduced the concept of premature cognitive commitment as an attitude that develops when a person first encounters a device in a particular context and that is then reinforced when the device is re-encountered in the same way. Langer identified several antecedent conditions that produce this type of attitude, including routine, repetition, and extremes of workload—all conditions that can occur in the automated cockpit. According to this reasoning, automation that is unchanging in its reliability (but less than 100% reliable) is more likely to induce a condition of complacency than automation that varies. That is, complacency is a function of the consistency of reliability of the automated task in a multitask environment.

Four main hypotheses were proposed. First, complacency should be potentially high for a group of subjects encountering automation with constant, unchanging reliability because this group is more likely to develop a premature cognitive commitment regarding the nature of the automation and its efficiency. On the other hand, subjects encountering inconsistent automation reliability should have a more open attitude concerning the efficiency of the automation and hence should be less likely to be complacent. We therefore predicted that the accuracy of operator detection of automation failure would be greater for variable-reliability than for constant-reliability automation of a system-monitoring task. Also, this effect should emerge with time spent under automation control, so that operator performance in the two conditions would diverge gradually from an initially equal starting point. A second hypothesis was that the initial level of automation reliability is important due to premature cognitive commitment; the higher the initial reliability, the higher the complacency potential. Third, an operator's trust in and reliance on an automated system tend to wane immediately after a failure (Moray & Lee, 1990; Riley, 1989). With several consecutive failures, trust should be minimal. We therefore reasoned that the effects of complacency on monitoring performance would dissipate following total automation failure, in which the automation failed to detect all instances of malfunctions in the systems-monitoring task. These three hypotheses were tested in the first experiment. Finally, given the failure to find evidence for complacency effects in Thackray and Touchstone's (1989) single-task study, and the anecdotal evidence that complacency occurs primarily when pilots are engaged in multiple duties (see Singh, et al.'s, in press, ASRS example), we proposed that all these predictions would hold only in a multitask environment in which the operator is responsible for more than one function. This hypothesis was tested in the second experiment.

EXPERIMENT 1

Method

Subjects

Twenty-four volunteers (10 men, 14 women) participated. Subjects ranged from 19 to 43 years of age, were right-handed, and had normal (20/20) or corrected-to-normal vision. Each subject was tested in two 2-hr sessions held over 2 days. They were paid $25 for completing the study. None of the subjects had any prior experience with the flight-simulation task used in the study.

Flight-Simulation Task

A revised version of the Multi-Attribute Task Battery (MAT) developed by Comstock and Arnegard (1992) was used. The MAT is a multitask flight-simulation package of the component tasks of tracking, system monitoring, fuel management, communications, and scheduling. The modified version that we developed allows each component task to be performed either manually or under automation control (R. Parasuraman, Bahri, & Molloy, 1991). In the present studies, only the monitoring, tracking, and fuel-management tasks were used, and only the monitoring task could be automated. The three tasks were displayed in separate windows of a 13-in. VGA color monitor.

System monitoring. The upper-left window in Figure 1 shows the system-monitoring task, which consisted of four vertical gauges with moving pointers and green "OK" and red "Warning" lights. The scales for the gauges were marked to indicate the temperature (TEMP1, TEMP2) and pressure (PRES1, PRES2) of the two aircraft engines. Normally, the green OK light was on and the pointers fluctuated around the center of the gauge within a fixed range in each direction from center. In each 10-min block of the simulation, 16 "system malfunctions"

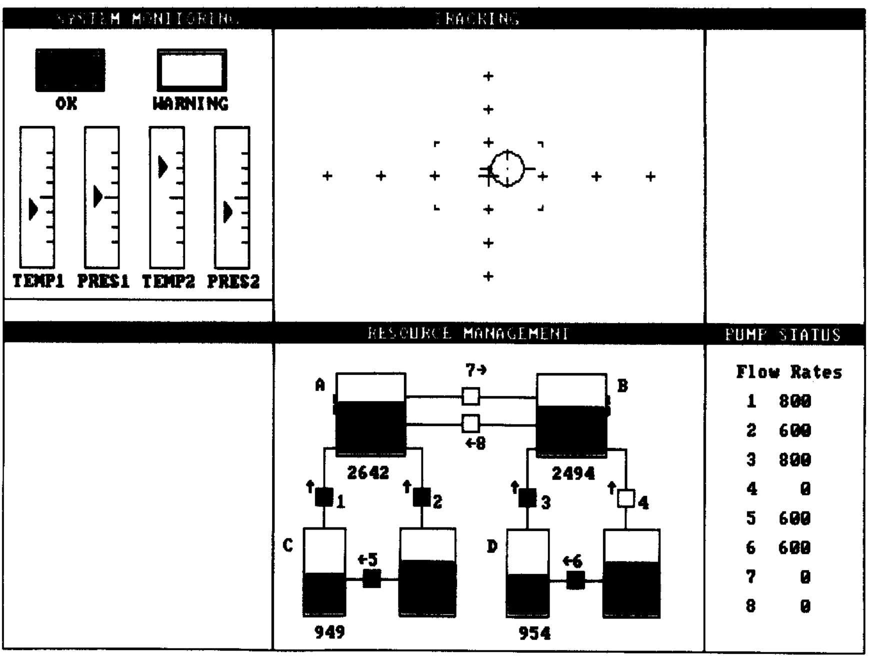

FIGURE 1 Revised MAT. From R. Parasuraman et al. (1991).

occurred at unpredictable intervals ranging from 13 to 72 sec. These occurred in a pseudorandom sequence except that an equal number appeared for each direction of pointer shift and for each vertical gauge. When a system malfunction occurred, the pointer on one of the four engine gauges went "off limits." That is, independently and at intervals according to a predefined script, the pointer shifted its center position away from the middle of the vertical gauge. These system malfunctions were normally detected and reset automatically—successful identification and correction of the malfunction being indicated by a red warning light coming on and then going off when the automation had corrected the problem in 4 sec. During this time, the subject's response keys were disabled to prevent manual input.

Under normal operating conditions, the automation routine detected and corrected the system malfunctions. However, from time to time the automation failed to detect a malfunction. The mean failure rates and the distribution of interfailure intervals are given later. Also, the automation failure was always a detection failure; the automation routine never made false alarms. When the automation failed, the operator was responsible for detecting pointer shifts occurring on any of the four gauges, regardless of direction, and responding by pressing one of the corresponding function keys (T1, T2, P1, or P2), which were labeled below each vertical gauge. When the automation routine failed, the green OK light remained on. Feedback was provided when the out-of-range status of a gauge was correctly identified, whether under automation or manually by the subject. 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. If the subject failed to detect a malfunction, the fault was automatically corrected 10 sec after the beginning of its occurrence.

If the subject responded appropriately to an automation failure by pressing the correct function key, the response was scored as a correct detection of automation failure. If the subject failed to detect the failure within 10 sec, the gauge was reset and the response was scored as a miss. Detection errors occurred when the operator detected an automation failure but incorrectly identified the gauge associated with the failure (e.g., the subject responded T1 for a malfunction in the temperature of Engine 2). All other responses were classified as false alarms. Thus, the performance measures for the system-monitoring task included (a) the probability of detection of automation failures, (b) reaction time (RT) for detection (to within a resolution of 0.1 sec), and (c) the number of false alarms and detection errors made.

Tracking. A first-order, two-dimensional compensatory tracking task with joystick control was presented in one window of the display (see Figure 1). Dashed X and Y axes were provided for reference. A green circular target symbol representing the deviation of the aircraft from its course fluctuated within the window in the X and Y directions according to a specified forcing function consisting of a sum of nonharmonic sine waves. The highest (cutoff) frequency of the forcing function was 0.06 Hz. Control inputs were provided by a displacement joystick using first-order or velocity control. If no control input was applied, the aircraft symbol drifted away from the center toward the edges of the window. The subject's task was to keep the aircraft within the central rectangle by applying the appropriate control inputs in the X and Y directions. The X and Y control inputs were sampled at 10 Hz to yield the X and Y deviations. Combined root-mean-square (RMS) errors were then computed for the samples obtained over each 1-sec period and averaged over a 10-min block to yield a mean RMS error score for a block.

Fuel management. This task was a simulation of the actions needed to manage the fuel system of the aircraft. Figure 1 displays the fuel (or resource) management window. The display for this task consisted of six rectangular regions that represented aircraft fuel tanks. Along the fuel lines that interconnected these tanks were fuel pumps capable of transferring fuel in one direction and at a specified flow rate. Subjects were required to maintain a specific fuel level within both of the main tanks by selectively activating pumps to keep pace with the fuel consumption in these main tanks. Tanks were depleted of fuel at a constant rate. Therefore, to maintain the task objective, subjects had to transfer fuel from the supply tanks using one or more of the eight fuel pumps. Pumps were activated by key presses; successive key presses toggled an associated pump on and off. A global measure of task performance was obtained by computing the mean RMS error in the fuel levels of tanks A and B (deviation from the required level of 2,500 gal). Fuel levels were sampled and RMS errors were computed for each 30-sec period; then they were averaged over a 10-min block to yield a mean RMS error score for a block.

Design

Automation Reliability was varied as a between-subjects factor (Constant or Variable Reliability) and Blocks (or Sessions) as a within-subject factor in a factorial design. Automation reliability was defined as the percentage of the 16 system malfunctions correctly detected by the automation routine in each 10-min block. Subjects were randomly assigned to two main groups: constant-reliability or variable-reliability automation. For the constant-reliability group, automation reliability was constant from block to block; for the variable-reliability group, it alternated between two values from block to block. There were twelve 10-min blocks in four 30-min sessions.

The effect of the overall level of automation reliability was examined by dividing the constant-reliability group into two subgroups. Automation reliability was chosen to be relatively high for half the subjects (87.5% or 14 of 16 malfunctions detected) and relatively low for the other half (56.25% or 9 of 16 malfunctions detected). Thus, for each 10-min block, subjects had to intervene for two automation failures if automation reliability was high and for seven automation failures if it was low. These levels remained constant across each 10-min block for each subgroup. In the variable-reliability condition, reliability alternated every 10 min from low (56.25%) to high (87.5%). The effect of the initial level of reliability was examined by dividing the variable-reliability group into two subgroups. Automation reliability alternated from low to high for half the subjects and from high to low for the other half. A different quasi-random distribution of interfailure intervals (IFIs) of the automation was used for each 10-min block. Although the mean automation failure rates and the distribution of IFIs differed for the low- and high-reliability subgroups, for the major comparison of interest (i.e., for the constant- and variable-reliability groups), the IFI distributions across all blocks were similar—both groups approximately describing a log normal form (see Figure 2).

Procedure

Each subject completed four 30-min sessions for a total of twelve 10-min blocks. First, subjects were given general instructions on the flight-simulation task. To ensure the equivalence of initial performance levels between the two automation-reliability groups, all subjects then received a 10-min practice block in which they performed all three tasks manually. Subjects were requested to give equal attention to all three tasks. At the end of the training session, feedback on performance was given for all three tasks. Following a 3-min break, the experimental phase began with subjects being informed that the monitoring task would be automated, so that they should focus their attention on the tracking and fuel-management tasks. However, they were informed that the automation routine was not perfectly reliable and that they had to supervise the automation and detect and respond to any malfunction that the automation failed to detect. No information concerning automation-reliability values or the distribution of IFIs was provided to subjects—other than the general instruction that the automation routine was not perfectly reliable. Performance feedback on only the tracking and fuel-management tasks was provided at the end of each 30-min session.

Subjects participated in the four 30-min sessions over a period of 2 days. They received a 10-min break between the first and second sessions on the first day. Subjects returned to the lab after 1 or 2 days. They were given a 10-min break between the third and fourth sessions. In each automation-reliability, condition, the

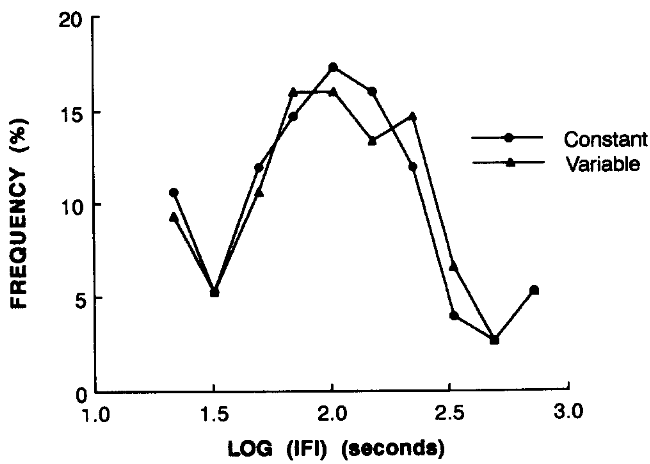

FIGURE 2 Frequency distribution of IFIs of automation for the constant-reliability and variable-reliability conditions.

automation routine was disabled for the last 20 min of the fourth session (Blocks 11 and 12) in order to simulate catastrophic automation failure (0% reliability). Thus, subjects were required to intervene manually all the time following the total failure. However, instructions for the last session were the same as for the first three sessions, during which subjects were required to intervene only occasionally.

Results

Practice

Performance levels during practice for the monitoring, tracking, and fuel-management tasks, each of which was performed manually, were compared for the constant-reliability and variable-reliability groups to examine whether the two groups differed in their initial ability to perform the tasks. There were no significant group differences on any of the five dependent measures used for the three tasks ($t < 1.08$, $p > .25$ in each case). For the monitoring task, the mean probability of detection of system malfunctions was .72, mean detection RT was 4.6 sec, and the mean number of false alarms emitted was 9.8.

System Monitoring

The efficiency of manual intervention in the automated monitoring task was assessed by computing the probability of detection of automation failures, the mean RT for detection, and the number of false alarms made. Detection errors (detecting an automation failure but incorrectly identifying the gauge associated with the failure) were rare and were not subjected to statistical analysis.

Detection rate of automation failures. Mean detection rates of automation failures are plotted in Figure 3 as a function of blocks. As Figure 3 shows, the probability of detection of automation failures was higher in the variable-reliability than in the constant-reliability condition, particularly in the later blocks. A 2 (Automation Reliability) × 12 (Block) analysis of variance (ANOVA) of the detection rate gave a significant effect for automation reliability, $F(1, 22) = 23.0$, $p < .0001$. In addition, the effect of blocks was significant, $F(11, 242) = 1.90, p < .05$, showing that detection rate improved over blocks. However, most of this improvement was contributed by the subjects in the variable-reliability condition, as indicated by a Group × Block interaction that approached significance, $F(11, 242) = 1.73$, $p < .07$. Tests of the simple effects of automation reliability were carried out to probe the main and interaction effects further. The effect of automation reliability was not significant in Block 1, $F(1, 40) = 2.67, p > .05$, but was significant for all subsequent 11 blocks, $F(1, 40) > 8.99$, $p < .01$. As Figure 2 indicates, the performance advantage for the variable-reliability group was present early and grew in size across the twelve 10-min blocks. Detection probability for the variable-reliability group quickly reached a high level, exceeding .8 after Block 4, but remained at a low level for the constant-reliability group. Following the total automation failure in Block 11, performance of the constant-reliability group

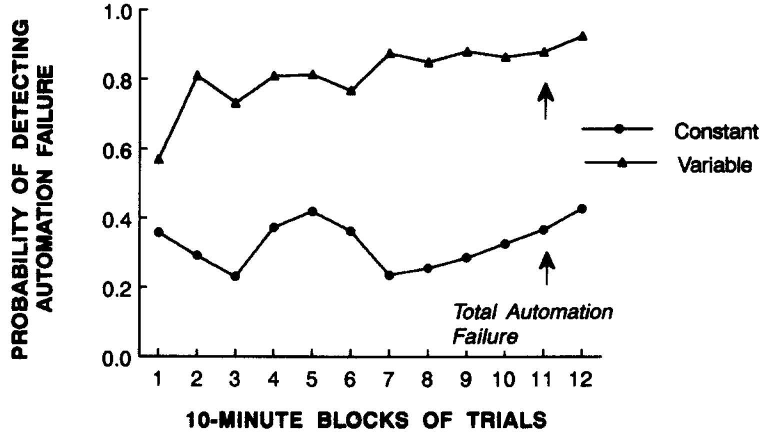

FIGURE 3 Effects of automation reliability (constant or variable) on probability of detecting automation failures during multitask performance.

improved but still did not recover to the level of the variable-reliability group.

Within the constant-reliability condition, half the subjects monitored a low-reliability system (57.25%), and the other half monitored a relatively high reliability system (87.5%). Furthermore, within the variable-reliability condition, automation reliability alternated from low to high for half the subjects and from high to low for the other half. To examine performance differences associated with the overall and initial levels of automation reliability, detection rates for these four subgroups were computed separately and are plotted in Figure 4. A 4 (Subgroup) × 12 (Block) ANOVA of these data gave significant effects for subgroup, $F(3, 20) = 7.21, p < .005$, and blocks, $F(11, 220) = 1.93, p < .05$. The group difference was primarily due to the difference between the constant-reliability and variable-reliability conditions, as described previously. Detection probabilities for the constant low-reliability ($M = .37$) and the constant high-reliability ($M = .28$) conditions were not significantly different. Similarly, there was no significant difference between the two subgroups of the variable-reliability condition (Ms $= .82$ and $.81$).

Effects of the location of an automation failure. When an automation failure occurred, it could involve any one of the four vertical gauges in the system-monitoring window. Because this window was in the upper-left corner of the display, it is possible that only automation failures involving the leftmost gauge were undetected by subjects in the constant-reliability condition because they were the furthest in the peripheral visual field. To examine this possibility, automation failure detection rates were partitioned by location and were analyzed in a 2 (Automation Reliability) × 4 (Failure Location) × 4 (Session) ANOVA. Once again, the main effects of automation reliability, $F(1, 10) = 9.48, p < .05$, and session, $F(3, 30) = 3.14, p < .05$, were significant, as was the Automation Reliability × Session

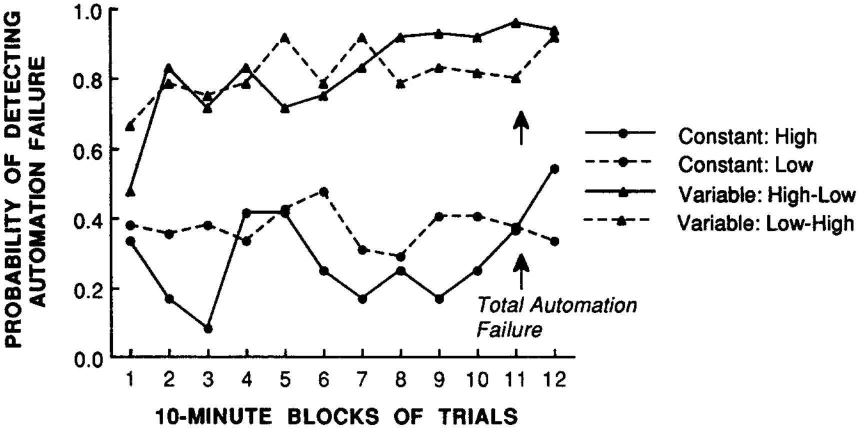

FIGURE 4 Effects of level (low or high) and variability of automation reliability (constant or variable) on probability of detecting automation failures.

interaction, $F(3, 30) = 2.98, p < .05$. There were no significant main or interaction effects involving Failure Location as a factor.

Combined-system (human–computer) detection rate. A combined-system measure of human–computer detection efficiency in the monitoring task was also analyzed. This was computed for each 10-min block by adding the number of malfunctions successfully detected by the automation routine to the number of malfunctions (automation failures) detected by each subject and dividing by the total number of malfunctions in each block (16). The combined-system detection rate is shown in Figure 5. System performance under variable-reliability automation was superior to that obtained under constant-reliability automation, $F(1, 22) = 12.16, p < .005$. As Figure 5 indicates, system performance was somewhat lower under constant-reliability automation in Blocks 1 to 10 but markedly lower following the total automation failure in Block 11. These trends were reflected in a significant effect of blocks, $F(11, 242) = 14.87, p < .0001$, and a significant Block × Group interaction, $F(11, 242) = 10.75, p < .0001$.

RT. RTs could not be computed for each of the twelve 10-min blocks separately because six subjects had 0% detection rates in at least one of the blocks. All six subjects were from the constant-reliability group. Hence, mean RTs were computed for each of the four sessions by averaging across the three successive blocks making up each session. Even so, the RT results should be interpreted with some caution due to the low numbers of trials comprising mean values. Mean RTs for the two automation-reliability conditions are shown in Table 1. A 2 (Automation Reliability) × 4 (Session) ANOVA showed that the effect of automation reliability was not significant, $F(1, 22) = 3.44$, $p > .07$. All other sources of variance were also not significant.

False alarms. False alarms were few in number, averaging a mean of 7.2 for the constant-reliability condition and 2.7 for the variable-reliability condition. A 2 (Automation Reliability) × 4 (Session) ANOVA gave no significant effects for any factor.

Tracking

Mean integrated RMS error on the tracking task is shown in Table 2. Both automation-reliability groups showed an improvement in tracking performance across sessions, $F(3, 66) = 18.34$, $p < .0001$. There was no significant group difference in tracking RMS error, and there was no significant Group × Session interaction.

Fuel Management

Performance on the fuel-management task is shown in Table 2. Mean RMS error decreased across sessions for both automation- reliability groups, $F(3, 66) = 5.87$, $p < .005$. There was neither a significant group difference in fuel-management RMS error nor a significant Group × Session interaction.

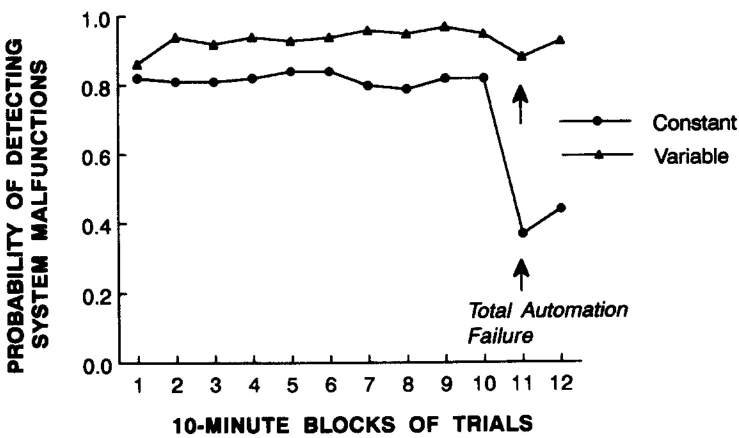

FIGURE 5 Effects of automation reliability on combined-system (human–computer) probability of detecting of automation system malfunctions.

TABLE 1
Mean RTs, in Seconds, of Detected Automation Failures for the Constant-Reliability and Variable-Reliability Conditions: Multitask Performance

	Condition	
Session	*Constant Reliability*	*Variable Reliability*
1	2.48 (0.95)	4.23 (1.02)
2	2.81 (0.89)	4.08 (0.96)
3	2.83 (1.42)	4.21 (1.16)
4	3.59 (1.27)	3.61 (1.13)

Note. Standard deviations are in parentheses.

TABLE 2
Mean RMS Error Scores for the Tracking and Fuel-Management Tasks for the Constant-Reliability and Variable-Reliability Conditions

	Condition			
	Constant Reliability		Variable Reliability	
Session	*Tracking*	*Fuel Management*	*Tracking*	*Fuel Management*
1	135.7 (54.9)	213.7 (219.9)	136.6 (64.6)	224.2 (122.5)
2	120.6 (46.0)	167.7 (163.3)	130.9 (60.6)	168.5 (133.4)
3	107.6 (46.5)	149.7 (161.4)	101.8 (43.4)	121.5 (76.7)
4	103.9 (40.4)	150.4 (164.8)	100.7 (44.2)	136.0 (73.9)

Note. Standard deviations are in parentheses.

EXPERIMENT 2

The aim of Experiment 2 was to examine whether the effects obtained in Experiment 1 could also be observed in a single-task environment. To be consistent with Experiment 1 and to ensure that the two automation-reliability groups were equivalent in initial performance levels, subjects trained and practiced on all three tasks. Following training, however, they performed only one task, system monitoring, which was under automation control. Automation reliability was varied as in Experiment 1.

Method

Subjects

Sixteen adults (11 men, 5 women), ages 19 to 44 years, participated. As in the first study, all subjects were right-handed, had normal (20/20) or corrected-to-normal vision, and were paid $25 for completing the study. Each subject was tested in two 2-hr sessions held over 2 days. None of the subjects had any prior experience with the flight-simulation task used in the study.

Flight-Simulation Task

The modified MAT task used in Experiment 1 was used here, except that the tracking and fuel-management windows were not active. All other features of the monitoring task were the same as in Experiment 1.

Design

The design of Experiment 2 was identical to that of Experiment 1—the only difference being that subjects performed only the monitoring task during the experimental phase. Automation reliability was either constant or variable and was either low or high—as in Experiment 1. Subjects were randomly assigned to either the constant-reliability condition or the variable-reliability condition, and within each condition automation reliability remained either low or high or instead varied from low to high (or vice versa) for the different subgroups, as in Experiment 1.

Procedure

The initial instructions and training were identical to those for Experiment 1. Following training, subjects performed all three tasks manually for a 10-min block. After a brief break, subjects performed only the system-monitoring task. They were told that their task was to detect failures of the automation routine that controlled the system-monitoring task. They were also informed that the automation routine was not completely reliable. All other aspects of the procedure were the same as those in Experiment 1.

Results

Practice

As in Experiment 1, performance on the three tasks during practice was analyzed to ensure that the two automation-reliability groups did not differ in their initial performance capabilities. There were no significant group effects on any of the five dependent measures of performance on the system-monitoring, tracking, and fuel-management tasks ($t < 1.47$, $p > .15$ in each case). For the monitoring task, mean detection probability was .68, mean detection RT was 4.8 sec, and mean number of false alarms was 4.1.

System Monitoring

The mean detection probabilities of automation failures for the two automation reliability groups are shown in Figure 6. Detection rates were uniformly high, and ANOVA gave no significant effects due to groups, sessions, or the Group × Session interaction. Mean RTs are shown in Table 3. Subjects responded significantly more quickly over sessions, $F(3, 42) = 10.734$, $p < .01$. Although Table 3 shows a tendency for RTs to be consistently higher in the constant-reliability condition than in the variable-reliability condition, neither this difference nor the Group × Session interaction was significant.

Very few false alarms were made (average = 1.3 for the constant-reliability condition, 1.4 for the variable-reliability conditions). Due to the low numbers, false alarms were not statistically analyzed.

DISCUSSION

The results indicate that performance consequences of automation-induced complacency can be detected in a multitask environment. Our first hypothesis was that the probability of operator detection of automation failure would be greater when the reliability of the automation routine controlling a system-monitoring task was

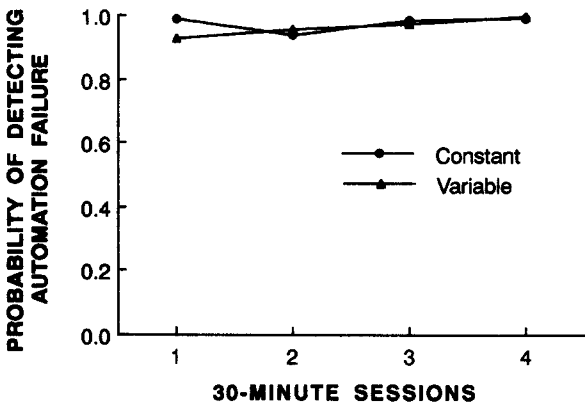

FIGURE 6 Effects of automation reliability on probability of detecting automation failures during single-task performance.

TABLE 3
Mean RTs, in Seconds, of Detected Automation Failures for the Constant-Reliability and Variable-Reliability Conditions: Single-Task Performance

	Condition	
Session	*Constant Reliability*	*Variable Reliability*
1	3.61 (1.45)	3.29 (0.83)
2	2.51 (0.61)	2.16 (0.84)
3	2.41 (0.87)	2.26 (0.53)
4	2.18 (0.59)	1.97 (0.54)

Note. Standard deviations are in parentheses.

variable rather than when it was constant over time. Also, this effect should emerge as time is spent under automation control. We predicted that initially the accuracy of manual intervention in the automated monitoring task would be roughly the same for the two automation-reliability conditions but that later the performance of the variable-reliability group would be superior to that of the constant-reliability group.

The results strongly supported these predictions. The detection probability of automation failure was markedly higher for the variable-reliability condition (M = .82) than for the constant-reliability condition (M = .33). The magnitude of the effect—a 149.4% difference in detection probability—is dramatic considering that, under single-task conditions, detection of system malfunctions was quite easy (detection probabilities averaged about .97; see Experiment 2). The mean detection rates also conceal the fact that, in the constant-reliability condition, some subjects simply failed to detect any automation failures in a given 10-min block; six subjects had at least one instance of 0% detection in a 10-min block, whereas no such instances occurred for subjects in the variable-reliability condition. Wiener (1981) proposed that the concept of complacency needed empirical support in order for it to be considered useful in the analysis of aviation safety. The present results suggest that such empirical evidence can be obtained, at least for the flight-simulation tasks used in the present study.

The effect of variations in the consistency of automation reliability emerged sooner than we anticipated, becoming present by the second block of the first session. Thackray and Touchstone (1989) suggested that long sessions might be needed to observe the performance consequences of complacency and that definitive evidence might require lengthy field studies with very infrequent automation failures. Partly in response to that suggestion, we tested subjects over four sessions, with the expectation that complacency effects might become apparent only in the last session. Surprisingly, although the effect of complacency on performance did increase with time, the effect was present in the very first session. Under the conditions of the present experiment, automation-induced complacency had performance consequences after only about 20 min spent under automation control.

Although the effects of complacency increased with time, there was no evidence of within-session, time-on-task effects on detection rate of automation failures—that is, there was no vigilance decrement (R. Parasuraman, 1986; Warm, 1984). Most studies reporting vigilance decrement have used relatively simple tasks, unlike the dynamic, complex display in the present study. Although not unequivocal, the evidence from previous studies of vigilance suggests that complex tasks such as the present one may not show decrement over time, although the overall level of performance may be low, as was the case in the constant-reliability automation condition (R. Parasuraman, 1986).

The second hypothesis—that the probability of operator detection of automation failure would be related to the initial level of automation reliability—was not supported. Detection rate did not differ for the low-reliability (57.25%) and high-reliability (87.5%) subgroups in the constant-reliability condition. Also, in the variable-reliability condition, detection rate did not differ for subjects who initially monitored under high reliability versus those who initially monitored under low reliability. This suggests that a high initial reliability, which may induce premature cognitive commitment (Langer, 1989), is not the only factor influencing complacency. Rather, the dominant factor influencing complacency seemed to be the consistency of performance of the automation. Subjects appeared to be lulled into complacency by unchanging automation, irrespective of the absolute level of reliability. Automation of variable reliability, on the other hand, may have provoked a more skeptical attitude (cf. Singh et al., in press) that promoted higher vigilance. Before concluding that the level of automation reliability had no impact, however, it should be noted that only two levels of automation reliability were used in the present study. Examination of a wider range of reliability levels might show a relation between automation-induced complacency and the level of automation reliability. It is possible that the "low" value of reliability used in the present study was not low enough to offset the complacency induced by the constant-reliability environment.

The third hypothesis—predicting recovery in performance for the constant-reliability group following total automation failure (i.e., 0% automation reliability)—

was only partially supported. Although detection of system malfunctions did improve following the catastrophic failure, detection rate did not recover to the level achieved by the variable-reliability group. It is possible that recovery might have been greater had we tested subjects for more than two blocks following the failure. If this were true, then it would suggest the intriguing conclusion that automation-induced complacency took only 20 min to develop but required greater time (> 30 min) to recover. In addition, recovery may have been greater if there were a consequence (i.e., impending crash or financial loss) for missing an automation failure. Palmer and Degani (1991) found that pilots using an automated checklist in simulated flight failed to detect automation failures when the failure did not affect flight safety. However, all four flight crews tested detected the checklist automation failure that would have led to a crash. This suggests that the saliency or aversiveness of failure consequences affects the rate of recovery following catastrophic automation failure.

Last, we found strong support for the prediction that the above three hypotheses would hold only under multitask conditions when the subject had responsibility for more than the function under automation control. When the subject simply had to "back up" the automation routine without other duties, variation in automation reliability had no effect on detection of automation failures, and monitoring was efficient. Detection rate was uniformly high (near 100%) in both automation-reliability conditions. The results suggest that automation-induced complacency is more easily detectable in a multitask environment when operators are responsible for many functions. This may account for the inability of Thackray and Touchstone (1989) to obtain reliable performance effects related to complacency in their study of simulated ATC, in which subjects performed a single task with and without an automated aid.

The single-task results also suggest a reconsideration of the relation between complacency and workload. Complacency is often linked to boredom arising from operator underload in highly automated systems—the implication being that boredom and complacency are related concepts. However, our finding that monitoring of automation failure is poor under multitask but not single-task conditions indicates that automation-induced complacency is not necessarily associated with low workload.

The proposal that complacency and boredom are distinct also raises the more fundamental question of the construct validity of complacency. Wiener (1981), in examining whether the term *complacency* was necessary, suggested the need for further research to answer the question. However, although hardly any relevant work has been done since that time, the term has been disseminated by virtue of its usage in aviation parlance and in ASRS reports. Complacency was operationally defined in our study as a failure to respond to an automation malfunction. So why use the term? Could the results be discussed simply in terms of the independent variables and previously established constructs such as vigilance and attention? As far as the present results are concerned, that is the case.

Yet, some arguments can be made in support of the view that the term *complacency* may still be needed. The distinction between complacency potential and performance is important: Only performance was examined in the present study. In that respect, the measure of performance and its underlying constructs, if any, can serve as descriptors—without need for the construct of complacency. However, it is not enough to say that complacent performance is the same as a failure of vigilance or attention, because consequences for performance may occur in other domains, not just in detection of automation failures as in the present study. Complacent performance may occur because human operators (a) failed to carry out some action that they believed had already been carried out by automation, or (b) they neglected to verify a checklist item due to the use of an electronic checklist (Palmer & Degani, 1991) or due to other possible behaviors. Thus, although complacency may not be necessary as a construct to describe performance changes in the presence of automation, it cannot be uniquely associated with any one performance-related construct such as boredom, vigilance, or workload.

Complacency potential may form a distinct attitude. In a preliminary construct validation of the Complacency Potential Rating Scale (CPRS; Singh et al., in press), complacency potential was found to be uncorrelated with attitudes toward general automation or automated decision aids (Igbaria & S. Parasuraman, 1991; S. Parasuraman, Singh, Molloy, & R. Parasuraman, 1992), with arousal, or with personality (Singh et al., in press). Our results also suggest that complacency is different from boredom or low workload. Clearly we know more about what complacency potential is not than what it is. Singh et al. (in press) found that CPRS items clustered around factors of Trust, Confidence, and Reliance on Automation. Each factor could have contributed to the present results. Because the present study was concerned primarily with investigating whether empirical evidence for performance effects of complacency could be obtained, it was not specifically designed to test competing hypotheses. Nevertheless, a plausible account of the results would suggest that the waxing and waning of trust with the success and failure of automated detection of malfunctions (Moray & Lee, 1990; Riley, 1989)—a *variable reinforcement schedule* in the terminology of learning theory—led to the superior performance of the variable-reliability group. Muir (1988) also suggested that a person's trust or distrust of a machine affects his or her use of automation, and this in turn affects the stability of that person's trust.

Other Contributing Factors

Some other factors that could have contributed to the performance advantage for the variable-reliability automation group should be considered. First, experimenter effects are unlikely to have produced the pattern of results obtained. All subjects received identical instructions: Subjects in the constant-reliability condition were not more or less encouraged to monitor the automated system than those in the variable-reliability condition. Second, the better monitoring performance exhibited by the variable-reliability group was not related to superior initial performance capability as such. Third, the higher detection rate of the variable-reliability group was not a function of speed–accuracy tradeoff. If subjects in the constant-reliability condition traded off accuracy for speed of detection of automation failure, then they should have responded faster to failure than the variable-reliability group, but this was not the case. Another possibility is that the poor monitoring performance of the constant-reliability group is related to the "signal rate" (i.e., automation failure rate), which affects detection performance (R. Parasuraman, 1986; Warm, 1984). However, there were no significant performance differences between the two subgroups of the constant-reliability condition—indicating that the absolute level of automation reliability, within the range studied here, did not affect the pattern of results.

Another possibility is that subjects in the constant-reliability condition may have fixated the monitoring window less often than those in the variable-reliability condition. If subjects fixated the tracking or fuel-management windows, the center of the monitoring window was about 5° away from fixation. Informal observation of subjects using a video camera did not reveal any systematic deviation in the pattern of eye movements in the two automation conditions, but we cannot rule out the possibility of small differences in scanning behavior between the two conditions. If subjects looked more often at the tracking and fuel-management windows or devoted more processing resources (Wickens, 1984) to these tasks—to the detriment of monitoring performance—one might expect performance on these tasks to be superior to that of the variable-reliability group. However, there were no significant differences in tracking or fuel-management performance between automation reliability conditions—demonstrating that the performance advantage of the variable-reliability group was not due to their transferring processing resources from tracking or fuel-management to system monitoring.

Practical Implications

The results of the present study are relevant to the debate on technology-centered versus human-centered approaches to the design of cockpit automation (e.g., S. Norman et al., 1988). The dominant tendency of the former approach has been to implement automation wherever possible in order to reduce pilot workload and to reap the benefits of economies such as fuel efficiency and reduced training costs. As a result, some commercial airline pilots have expressed the opinion that their role is being reduced from active control to system monitoring—a role to which they are ill-suited (James, McClumpha, Green, Wilson, & Belyavin, 1991). Several researchers have also discussed the potential adverse effects of high levels of automation on crew monitoring efficiency (Chambers & Nagel, 1985; R. Parasuraman, 1987; Wiener, 1987). The contention has often been put forward that taking the pilot "out of the loop" by automating a manual function degrades system awareness and manual skills, so that the pilot cannot intervene effectively if the automation fails. In fact, the evidence in support of this view is by no means consistent (R. Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1990). However, the present results do support this position for the system-monitoring aspect of performance.

That monitoring was more efficient with variable-reliability automation could suggest that inserting artificial automation failures at random intervals might serve to reduce operator complacency. This has been examined before with sonar displays and industrial inspection (e.g., Murrell, 1975; Wilkinson, 1964; see also Rizy, 1972). An automated system that must be monitored could be designed to simulate a variable-reliability system by including (at variable intervals) artificial failures that would require an operator response (and that would have some performance consequences). However, any possible benefit of such a system is likely to be outweighed by several disadvantages. First, artificial signals may not result in a real improvement in monitoring efficiency (i.e., detection sensitivity or d') because, although increasing the detection rate, they also raise the operator false alarm rate (Murrell, 1975). This might result in system false alarms that increase operator false alarms. Second, any such system is unlikely to be appreciated or trusted—and therefore used—by pilots or other operators of automated systems. Artificial signals may also lead to unacceptable momentary increases in workload if they occur at critical phases of flight. A more feasible solution might be to use adaptive function allocation (R. Parasuraman et al., 1990; Rouse, 1988) to transfer manual control of an automated task temporarily to the operator during noncritical work periods. R. Parasuraman, Mouloua, Molloy, and Hilburn (1992) found that brief periods of adaptive manual control led to subsequent improvement in monitoring efficiency under automation control.

An obvious criticism of this study concerns the values used for automation reliability, which are clearly artificial. A "high" reliability of 87.5% would be unacceptable in any real system, but we had to use such values due to time constraints. Using automation reliability values of 99% or higher would require subjects to be tested over very many sessions in order to gather reliable data. It would be valuable to replicate the present results using reliability levels closer to those typical of modern automation technology, but this could prove time-consuming and expensive.

The high (but not perfect) reliability of automated systems raises an issue: Does the fact that human monitoring is inefficient matter, given that automated monitoring can be near-perfect? Our results suggest that it does matter when conditions likely to produce operator complacency are present, even if machine monitoring is very reliable. Many studies of joint human–computer performance have found that aided performance is generally better than human performance or computer performance alone (Corcoran, Dennett, & Carpenter, 1972; R. Parasuraman, 1987; Sorkin & Woods, 1985). For example, in their study of simulated ATC, Thackray and Touchstone (1989) found that combined-system (i.e., human–computer) monitoring performance was 99% reliable in detecting aircraft-to-aircraft conflicts, whereas unaided humans detected only 89% of such conflicts. However, in their study, subjects did not exhibit complacency, and monitoring was very efficient. The present results suggest that, when complacency is likely to be present, overall system performance may be poor. Combined-system performance was very efficient under the variable-reliability condition, but less efficient under the constant-reliability condition that promoted complacency, particularly following total automation failure. This result suggests that operator complacency could limit system performance following catastrophic automation failure even if the automation is very reliable (e.g., between 99% and 100%). Thus, high automation reliability ensures high system reliability except when the automation fails (which it will do at some time), in which case system performance drops precipitously.

The present results suggest that automated monitoring does not guarantee reliable system performance due to the potential for human monitoring to exhibit the consequences of automation-induced complacency. Thus, as Palmer and Degani (1991) pointed out, automated monitoring does not necessarily provide true redundancy to human monitoring because the presence of automation affects the efficiency of human monitoring. In the present study, we found that the consistency of the reliability of the automated monitoring system influenced the efficiency of human monitoring. Many other automation characteristics may also affect complacency. Palmer and Degani (1991) also reported complacency-type effects in a study of the effects of an electronic checklist on system monitoring during a line-oriented flight simulation. Pilots used a conventional paper checklist, an electronic checklist requiring pilot acknowledgment of each item, and a fully automated checklist that did not require manual acknowledgment. Palmer and Degani reasoned that pilots would trust the fully automated checklist and would not perform their own manual check of system state. To test this hypothesis, "configuration probes"—aircraft subsystems that were set to an incorrect state (e.g., the anti-skid system off during descent)—were inserted at different points during the simulated flight. These items were wrongly shown as correct by the electronic checklist. Palmer and Degani found that pilots missed several probes (6.5 of 12) when using either version of the electronic checklist but missed only 1 of 12 items when using the conventional paper checklist. Palmer and Degani concluded that both electronic checklist designs led pilots to not conduct their own checks of the relevant system variables.

The problem of overreliance on automation is known to the aviation industry. A recent large-sample survey of attitudes toward cockpit automation among British pilots revealed that a majority believed that pilots of automated aircraft rely too heavily on automation (James et al., 1991). Techniques to counter pilot overreliance on automation, particularly under high-workload conditions, have also been a concern regarding the Airbus A310 and A320, which are the most highly automated commercial aircraft currently flying (Speyer, Monteil, Blomberg, & Fouillot, 1990). Pilots training to fly the Airbus A320 are exhorted to avoid overreliance on and overconfidence in the aircraft's capabilities (Stix, 1991). Recently it was reported (National Public Radio, 1992) that, in the wake of the January 1992 crash of an Air Inter Airbus A320 in Eastern France, French airlines are enforcing a policy that pilots should intermittently take manual control of automatic systems. Such measures, as well as those based on adaptive function allocation (R. Parasuraman et al., 1992), may provide possible countermeasures to automation-induced complacency.

ACKNOWLEDGMENTS

Part of this article was presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH, April 1991.

The research was supported by Contracts NAG-17 (University of Minnesota subcontract) and NAG-1-1296 from the NASA Langley Research Center, Hampton, VA. Alan Pope was the technical monitor. The views presented in this article are those of the authors and do not necessarily represent the views of NASA.

We thank Jonathan Gluckman, Peter Hancock, Brian Hilburn, Patricia May, Alan Pope, Joel Warm, and two anonymous referees for their comments on this study.

REFERENCES

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.

Chambers, N., & Nagel, D. C. (1985). Pilots of the future: Human or computer? *Communications of the ACM, 28,* 1187–1199.

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.

Corcoran, D. W. J., Dennett, J. L., & Carpenter, A. (1972). Cooperation of listener and computer in a recognition task: II. Effects of computer reliability and "dependent" versus "independent" decisions. *Journal of the Acoustical Society of America, 52,* 1613–1619.

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

Igbaria, M., & Parasuraman, S. (1991). Attitudes toward microcomputers: Development and construct validation of a measure. *International Journal of Man–Machine Studies, 35,* 553–573.

James, M., McClumpha, A., Green, R., Wilson, P., & Belyavin, A. (1991, March). Pilot attitudes to flight deck automation. In *Proceedings of the 33rd Annual Meeting of the Human Factors Society* (pp. 107–111). Denver: Human Factors Society.

Kingsbury, J. A. (1986). *Air traffic control automation: An AERA for this century* (Tech. Rep. No. MP-86W28). McLean, VA: MITRE Corporation.

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

Moray, N., & Lee, J. (1990). *Trust and allocation of function in the control of automatic systems* (EPRL Tech. Rep. No. 05). Urbana–Champaign: University of Illinois, Department of Mechanical and Industrial Engineering.

Muir, B. M. (1988). Trust between humans and machines, and the design of decision aids. In E. Hollnagel, G. Mancini, & D. D. Woods (Eds.), *Cognitive engineering in complex dynamic worlds* (pp. 71–83). London: Academic.

Murrell, G. A. (1975). A re-appraisal of artificial signals as an aid to a visual monitoring task. *Ergonomics, 18,* 693–700.

National Public Radio. (1992). *February 25 broadcast of "Morning Edition"* [Transcript]. Washington DC: Author.

Norman, D. A. (1991). Cognitive science in the cockpit. *CSERIAC Gateway, 2*(2), 1–6.

Norman, S., Billings, C. E., Nagel, D., Palmer, E., Wiener, E. L., & Woods, D. D. (1988). *Aircraft automation philosophy: A source document.* Moffett Field, CA: NASA Ames Research Center.

Palmer, E., & Degani, A. (1991, April). *Electronic checklists: Evaluation of two levels of automation.* Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 43.1–43.39). New York: Wiley.

Parasuraman, R. (1987). Human–computer monitoring. *Human Factors, 29,* 695–706.

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems* (Tech. Rep. No. CSL-N90-1). Washington, DC: Catholic University of America, Cognitive Science Laboratory.

Parasuraman, R., Bahri, T., & Molloy, R. (1991). *Adaptive automation and human performance: I. Multi-task performance characteristics* (Tech. Rep. No. CSL–N91–1). Washington, DC: Catholic University of America, Cognitive Science Laboratory.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1992). *Training and adaptive automation: II. Adaptive manual training* (Tech. Rep. No. CSL–N92–2). Washington, DC: Catholic University of America, Cognitive Science Laboratory.

Parasuraman, S., Singh, I. L., Molloy, R., & Parasuraman, R. (1992). Automation-related complacency: A source of vulnerability in contemporary organizations. In R. Aiken (Ed.), *Information processing 92* (Vol. 2, pp. 426–432). Amsterdam: Elsevier Science.

Riley, V. (1989). A general model of mixed-initiative human-machine systems. In *Proceedings of the 33rd Annual Meeting of the Human Factors Society* (pp. 124–128). Denver: Human Factors Society.

Rizy, E. F. (1972). *Effect of decision parameters on a detection/localization paradigm quantifying sonar operator performance* (Tech. Rep. No. R-1156). Arlington, VA: Office of Naval Research, Engineering Psychology Research Program.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*, 431–443.

Singh, I. L., Molloy, R., & Parasuraman, R. (in press). Automation-induced "complacency": Development of the Complacency Potential Rating Scale. *The International Journal of Aviation Psychology.*

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human–Computer Interaction, 1*, 49–75.

Speyer, J. J., Monteil, C., Blomberg, R. D., & Fouillot, J. P. (1990). Impact of new technology on operational interface: From design aims to flight evaluation and measurement. In *Aircraft trajectories: Computation—prediction—control. Volume 1* (AGARDograph No. 301, pp. 11.1–11.37). Neuilly sur Seine, France: North Atlantic Treaty Organization.

Stix, G. (1991). Along for the ride? *Scientific American, 265*, 95–106.

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

Warm, J. S. (Ed.) (1984). *Sustained attention in human performance*. London: Wiley.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Orlando, FL: Academic.

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. (1987). Application of vigilance research: Rare, medium, or well done? *Human Factors, 29*, 725–736.

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

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

Wilkinson, R. T. (1964). Artificial "signals" as an aid to an inspection task. *Ergonomics, 7*, 63–72.