Journal of Information Technology Impact
Vol. 1, No. 3, pp. 119-143, 1999

Alerting System Assertiveness, Knowledge, and Over-Reliance

Stuart D. Davis* and Amy R. Pritchett**
Georgia Institute of Technology
Georgia, U.S.A.

Abstract

Alerting systems have become an integral part of many complex technical systems, including aircraft cockpits. A simulator study examined the potential benefits of presenting system failure information to pilots using several levels of alerting system knowledge and assertiveness. The results indicate that the subjects used the alerting system more as an attention-directing signal than as a diagnostic tool. The results also indicated that increasing usefulness of information to the subjects seemed to stop at a limited amount of information, and that additional information provided beyond that point may actually slow the decision making process. An over-reliance test suggests that when an alerting system gives erroneous information that conflicts with other cockpit indications, serious mistakes can be made and pilots may have trouble correctly resolving the conflicts.

Keywords: alerting systems, cockpits, over-reliance.

Introduction and Motivation

Alerting systems have become an integral part of many complex technical systems, including aircraft cockpits, power plant control rooms, and medical operating rooms. They are typically introduced to monitor a process or a set of parameters and alert the operator when certain criteria are violated. Historically, alerting systems were used as attention-directing mechanisms to draw an operator's attention to a parameter that was outside of pre-defined limits. These early alerting systems were based on the same data that was available to the operator, and could therefore be thought of as the automated equivalent of red lines painted on the face of the gauges, indicating when a parameter exceeded its limitations. These alerting systems simply "alerted" the operator to a problem and the operator maintained full authority to react to the indication.

While these basic systems are still in use today, newer alerting systems have evolved into sophisticated information processing systems, capable of performing several functions: hazard detection; attention-directing; diagnosis and display of status; and presentation of guidance or command information. The hazard detection process involves monitoring the states of the process and the environment, and making a binary decision that a hazardous condition exists when the hazard level predicted from the states exceeds a predefined threshold. This detection may be quite simple – monitoring of a single sensor for violation of pre-defined threshold, for example – or it may be quite involved. For example, the information from many sensors may be integrated to develop a prediction of a hazard expected in the future, using computation-rich algorithms and models of system behavior. More than one type of hazard may be protected against, requiring the alerting system to have different 'modes' of behavior; for example, the Enhanced Ground Proximity Warning System in air transport aircraft may generate alerts based on a number of different hazard metrics, such as proximity to terrain in general, a comparison on proximity to terrain compared to the readiness of the aircraft to land as measured by flap extension, gear extension and position relative to an airport, or projected flight towards terrain regardless of proximity, Finally, hazard detection may have several different 'levels', such as precautionary alerts when a hazard is possible, 'cautions' when a hazard is likely, and 'warnings' when a hazard requires an immediate response.

When a hazard has been detected, an attention-directing signal or 'alert' is presented to the operator. Additional status information may also be presented. For example, a traditional alerting system generated an aural cue to notify the operator of a problem. More recent alerting systems can present a variety of information to the operator. For example, increasing the saliency of information about the hazard on visual displays may be used as a precautionary presentation of information. Once the hazard is assessed to be quite likely, a wide-range of information may be presented, including descriptions of system status and diagnoses of the situation, a capability very similar to that provided by expert-systems and decision-aids.

Beyond providing information about the current situation, alerting systems may also command resolutions. This command information may be generated actively by the system, or it may be procedurally associated with the generation of an alert, such as requiring a pilot to shut down an engine when presented with an engine fire light. This guidance and command information is generally reserved for time-critical warnings in which the operator must act with a minimum of delay.

Taking a broader definition of an alerting system that captures all of the functions currently instantiated in their sophisticated capabilities, an alerting system's role in the system is to provide the operator with key elements of information at critical times in the proper format. Designing an alerting system to perform this role in a manner beneficial to the operator can be difficult given the conditions for which these systems are intended: onset of mission-critical or life-critical hazards which require quick and accurate resolutions. This presents the system designer with the design challenges of pre-determining what information the operator needs, and how to provide it. This may be viewed as defining the *role* of the alerting system, including the detail of the information it provides, and the amount of assertiveness it has in interacting with the operator, such as quietly presenting information or noisily commanding a resolution.

In order to create an alerting system effective in helping an operator resolve hazards, it is useful to consider the types of processes the operator may follow. The exact process used in any one instance is chosen opportunistically by the operator in response to their capabilities and environmental demands. However, two models of decision making represent extremes in how an operator may respond to a hazard, and a brief discussion of both can illustrate the range of potential information needs and alerting system roles.

The most thorough and time-consuming decision process is modeled by Rasmussen's decision ladder (1994). This model describes the sequence of basic information processes that may occur in formal decision making, shown schematically in Figure 1. These steps illustrate the processes used in difficult problem solving that involves a high degree of uncertainty. This method assumes the operator will actively follow several steps in sequence, including detecting the need for action, information collection, situation analysis, diagnosis, goal review, option evaluation, planning a response, and then executing it. Beyond the traditional role of detection, sophisticated alerting systems have the potential to provide the information required at each step in their decision, with either a supporting role (providing background information) or an active role (completing specific steps for the operator and presenting the results to him or her).

For the operator to follow this extensive, complete decision process has the benefit that it engenders thoughtful, well-considered resolutions to the hazard. However, many hazards demand a very quick response from the operator. While an alerting system may help the operator more quickly generate a well-thought-out response by providing the right information at the right time, the sheer amount of information to be presented, interpreted and evaluated may force the operator to take short-cuts during their decision making process. These short-cuts can occur at any stage of the decision-process; for example, while observing the environment, the operator may recognize a situation for which they already know the task or procedure they should use to resolve it, allowing them to 'skip' the stages of identifying the problem, predicting consequences, etc.

In situations where the operator is extensively using these heuristic short-cuts, their decision process may be better described by the recognition primed decision making models. In recognition primed decision making, the operator makes decisions by recognizing patterns (cues) in the environment that match previously learned situations. The operator recognizes the goals, expectations, and actions required without performing extensive diagnosis or evaluation (Klein, 1998). The reaction to these situations has been learned through training or experience; for example, pilots receive extensive training in responding to likely hazards in order to enable immediate reactions. For these conditions, the alerting system may only be required to provide the operator with the key elements of information he needs to recognize critical situations. This may be in the form of a simple attention-directing mechanism to focus the operator's attention, or it may be highlighting a specific pattern of parameters that indicate a critical malfunction. The requirements for each situation may again vary, depending on the training of the operator, the criticality of the decision, and the time available in which to make the decision.

Just as it is difficult to predict which type of decision-making process the operator *will* follow, it is also difficult to pre-determine which process the operator *should* follow, and consequently which process the alerting system should be designed to support. As noted before, decisions made following the full decision-ladder tend to be thorough and accurate, at the risk of requiring more time than available to resolve a hazard. Conversely, recognition-primed decision making can be very quick, but is sensitive to misinterpretations of the situation leading to erroneous actions, and requires training for every individual situation which may need to be identified. This trade-off is reflected in alerting system design as well. The amount of information required from the alerting system to support the operator in following an extensive decision process may be overwhelming; however, if the alerting system only provides limited cues and/or commands, they must be taken at face value by the operator.

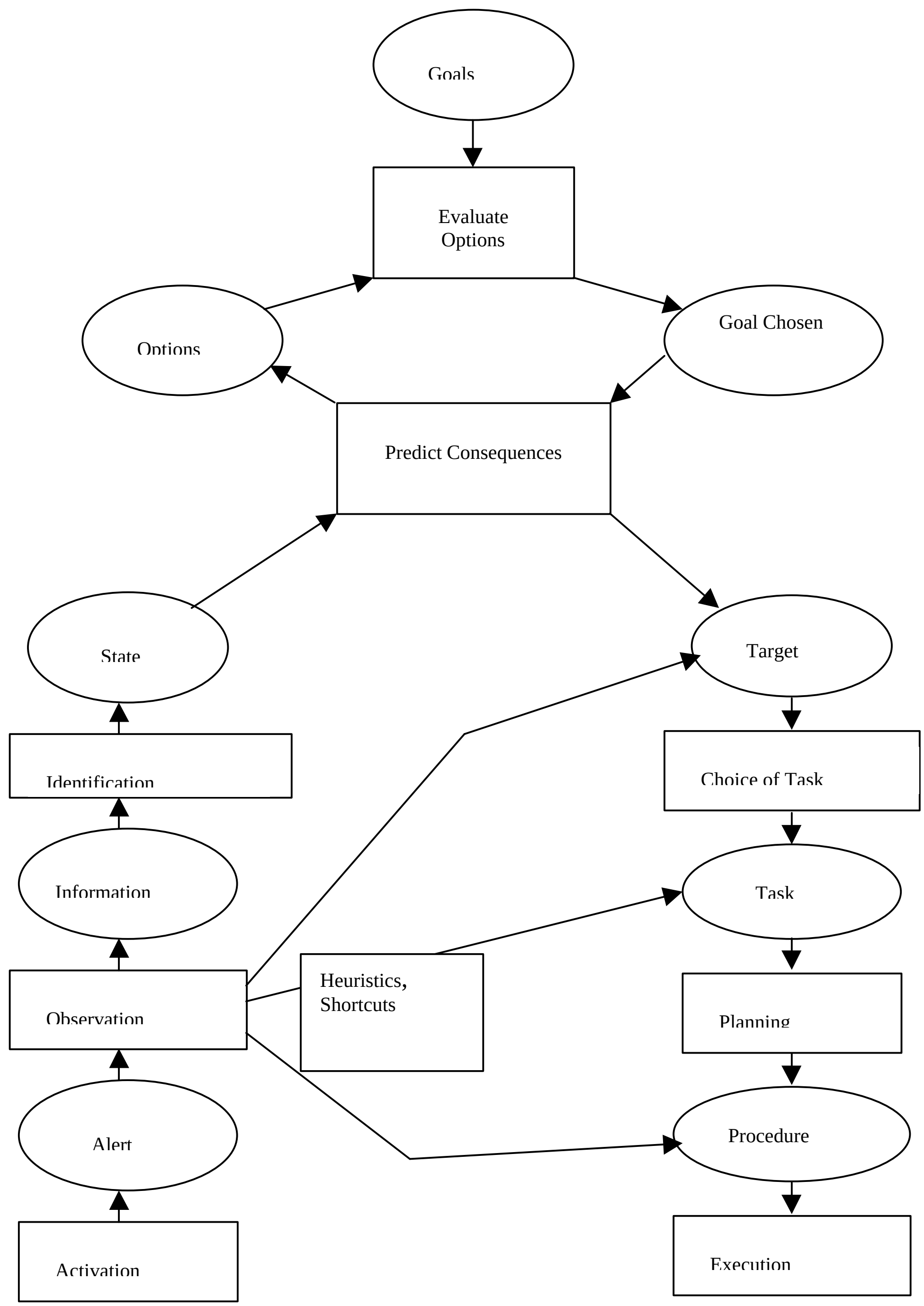


Figure 1. Rasmussen's Decision Ladder

The method of decision making also affects the role of the alerting system, in terms of its assertiveness. A passive system with no attention-directing mechanism could present information in a manner intended to support whatever process the operator may choose to follow, while an aggressive system may issue aural, written, and even tactile commands directing the operator on what his next actions should be to resolve a problem. Beyond the benefit of directing attention to hazards that might not be noticed in time by the operator, existing research provides conflicting guidance about the benefits of each role, from pilots performing with greater accuracy when provided with assistance (Prevot and Onken, 1996) and accepting the assistance of an alerting system that provides explanatory information (DeCelles, 1991; Hicks and Ross, 1990), to automation increasing cognitive workload (Endsley, 1997).

Another related concern in the design of alerting systems is the potential for the operator to over-rely on the alerting system. While alerting systems are becoming increasingly reliable, it is still possible for them to give faulty guidance to the operator. This fallibility may result from mechanical weaknesses in the sensors driving the alerting system, or from the probabilistic nature of the judgements alerting systems must make based on uncertain information. Recent research has shown that reliance on automation is greatly affected by several variables, such as trust in the automation system's reliability (Lee and Moray, 1992; Deaton and Glenn, 1999; Mosier, 1997). Preventing over-reliance on an alerting system requires the operator being able to understand whether the alerting system's output is appropriate for the situation. An operator following an extensive decision process is less likely to become over-reliant on an alerting system if he or she is performing deliberate analysis on the information. On the other hand, an operator using recognition primed decision making is more susceptible to over-reliance if he is depending on highly salient cues from an alerting system to make diagnoses (Mosier, *et al*, 1997).

This study was conducted to specifically examine, through a flight simulation experiment of new cockpit alerting systems, the following questions:

1. Information. How much information should an alerting system provide to an operator? What are the key elements of information the operator needs to meet his decision making requirements in his particular circumstances?

2. Assertiveness. What level of assertiveness should an alerting system have? Should the system simply present information to the pilot, alert him when the system determines that a problem exists, or provide advice or directives on how to recover from a failure? Are higher levels of assertiveness justified considering the possible detrimental effects?

3. Over-reliance. In conditions where the alerting system provides correct guidance to the operator, it is desired for the operator to use the alerting system to its fullest and react quickly. However, do these quick reactions come at the expense of the operator failing to recognize erroneous alerting system guidance?

Experiment Objectives

The objective of this experiment is to determine if alerting system knowledge and assertiveness affect pilot usage in diagnosing system failures. The experiment examined the following issues:

1. Ascertain how the level of knowledge of the alerting system affects pilot ability to diagnose system failures.

2. Ascertain how the level of assertiveness of the alerting system affects pilot ability to diagnose system failures.

3. Examine how pilots will respond to alerting system commands that are not supported by – or erroneous and therefore in conflict with – other cockpit indications.

Experiment Design

Overview

A simulator evaluation was conducted using the US Army's UH-60 Simulated Flight Training System (SFTS) at Fort Rucker, Alabama. The primary experiment tested pilots on their use of an alerting system with varying levels of information concerning system malfunctions and varying levels of directive assertiveness. A second experiment tested pilot over-reliance on an alerting system despite conflicting cockpit indications.

<u>Simulator Setup</u>

The UH-60 SFTS is a full motion simulator with almost all of the system functionalities of the actual aircraft. The simulator operator can input system failures at specified intervals through a touch screen interface in the rear of the cockpit. The aircraft cockpit warning system consists of a caution/advisory panel with 72 capsule lights, a master caution panel, and an audio warning system. This system was used in its normal mode to alert subjects to applicable system failures.

An additional warning system was added to the cockpit to allow the experimenter to provide the subjects with varying levels of additional information on the status of the aircraft and any current malfunctions. This warning system consisted of a laptop computer connected to a flat panel LCD screen. The laptop computer was operated from the simulator operator's position in the rear of the simulator and the screen was positioned in the cockpit in the center of the windscreen above the glare-shield. A speaker system was also connected to the computer to allow for auditory alerts in some test conditions. The additional information was presented as text messages projected onto the LCD screen. This additional display system was referred to as the Prototype Alerting and Display System, or "PADS". The PADS system provided system failure malfunction indications to the pilot through a textual readout on the PADS display as well as through normal cockpit indications (i.e., fuel gauge or oil pressure gauge).

Subjects

The subjects were twelve active duty Army helicopter pilots. The pilots were all qualified in the UH-60 helicopter and had between two and twenty years of operational helicopter flying experience. Total flight time ranged from 440 hours to 6800 hours, and UH-60 time ranged from 23 hours to 2500 hours.

Main Experiment

The main experiment was designed to test the effect of alerting system knowledge level and assertiveness level on pilots' ability to diagnose system failures. It consisted of twelve separate scenarios for each subject.

<u>Independent Variables</u>

The first experiment was designed as a two-factor experiment. These factors were the knowledge level of the system and the assertiveness of the system.

<u>System Knowledge</u>. The ability of the new warning system to diagnose system malfunctions was divided into six levels of knowledge. These levels of knowledge determined how much information the system provided to the pilot on the status of the aircraft. The levels of system knowledge are:

1. System diagnostics: General
2. System diagnostics: Some detail
3. System diagnostics: Detailed
4. Detailed system diagnostics and system implications
5. Detailed system diagnostics and aircraft implications
6. Diagnostics, implications, and recovery instructions: Recommendation or command

The amount of information provided to the pilot increased at each level. Examples of the information provided to the pilots at the lowest and highest amounts of knowledge for a sample malfunction (dual generator failure) are shown in Figures 2 and 3. The information provided at the levels between are scaled accordingly.

- Generator failure

Figure 2. Example of Information Provided the at "System Diagnostics: General" Level

- TURN SAS 1 OFF.
- RESET #1 AND #2 GENERATOR SWITCHES.
- TURN SAS 1 ON.
- #1 & #2 Generators failed.
- No AC power.
- Aircraft operating on battery power only.
- Generators can be reset.
- Aircraft can continue flight on battery power for 20 minutes in a degraded mode.

Figure 3. Example of Information Provided the at
"Diagnostics, implications, and commands for recovery" Level

System Assertiveness. The two levels of system assertiveness were informing/recommending and alerting/commanding. The subjects experienced only one level of assertiveness during their experiment. The amount of information provided to the subjects was the same at both levels of assertiveness. The difference in the two levels was in the provision of an audible alert for one level and in the provision of a recommendation versus a command at the highest knowledge level.

An audible alert was provided in the alerting/commanding command mode coincident with the display of the textual information. No audible alert was provided in the informing/ recommending mode. At the highest level of system knowledge, the informing/recommending system made a recommendation to the pilot on the best action to take in response to the existing malfunction. At the same knowledge level, the alerting/commanding system commanded the pilot to perform an action to correct for a system failure.

As part of the initial briefing, the subjects using the informing/recommending mode were told that the new alerting system could be used as an additional source of information in the cockpit, but that all information should be verified with other indications. The subjects using the alerting/commanding mode were told that the new alerting system could be used as the primary source of information concerning aircraft malfunctions. The intent was to make the alerting/commanding mode of the system more assertive and a more salient information source. Some of the resulting differences in usage may be due to the briefing effect from these initial instructions.

Test Matrix. Using these different knowledge and assertiveness levels, a fully-populated two-factor experiment was conducted. The knowledge levels were tested within subjects; the assertiveness levels were tested between subjects. Each subject completed a total of twelve runs, with two runs at each of the knowledge levels; each subject experienced only one assertiveness level, so that he or she saw PADS act consistently throughout the experiment.

Scenarios

For the primary experiment, twelve system failure scenarios were presented to each subject. The twelve scenarios were presented in a random order to each subject. Each scenario presented the pilot with a different system malfunction or failure that required him to take some action and then make a decision regarding the completion of the mission. The malfunctions required accurate diagnosis and reaction in a short period of time to maintain the safety of the aircraft.

The scenarios were all intended to exact a tactical response to a situation and an immediate determination of the consequences for the mission.

The system failures were introduced to the subjects simultaneously with normal cockpit systems and with the PADS added for the experiment. The additional display provided textual information to the pilots using one of the twelve levels of system knowledge and authority as described above. The list of scenarios is shown in Table 1.

TABLE 1. Malfunction Scenarios

Scenario #1: Hydraulic failure
Scenario #2: % Torque split
Scenario #3: Generator failure
Scenario #4: Engine failure
Scenario #5: Engine oil temperature
Scenario #6: Fuel pressure loss
Scenario #7: Rotor vibration
Scenario #8: Engine high-speed shaft failure
Scenario #9: Engine fire light illuminated with no fire
Scenario #10: Airspeed indications incorrect
Scenario #11: Crack in tail rotor spar
Scenario #12: Main transmission oil pressure slowly decaying

The first six scenarios may be considered “trained”. These are malfunctions that the subjects are trained on more regularly during their annual simulator training exercises and they are therefore more familiar with the cockpit indications for these malfunctions. The second six scenarios (7-12) may be considered as “less-trained”, as these malfunctions are not trained as regularly in the simulator, and the subjects are less familiar with their indications. One reason for the difference is that the "trained" malfunctions are more easily identified with cockpit indications, so they are good candidates for simulation training.

Over-reliance Experiment

Two additional scenarios were added to the end of the experiment (scenarios 13 and 14) to test the tendency of the pilots to trust the new alerting system after using it for only a short time. These scenarios followed immediately after the twelve scenarios in the main experiment. The subjects were not aware that these two scenarios were part of a separate experiment.

The independent variable for the over-reliance experiment was the reliability of the alerting system, created through scenario design. Two scenarios were presented to each subject for this experiment. The knowledge levels (and the PADS displays) were identical for each subject.

The first scenario (#13) provided the pilot with information that was not available from the aircraft's inherent warning system. The malfunction shown on the PADS display was an impending main rotor blade failure with catastrophic results. There was no other information available to the pilots to confirm or disconfirm this impending malfunction. The information on the display was correct, and if followed, prevented an incident. The intent was to further increase the subjects' trust in the system.

Scenario #14 attempted to examine how pilots would respond to alerting system commands that were not supported by – or were erroneous and therefore conflicted with – other cockpit indications. For this experiment, the PADS incorrectly diagnosed the malfunction and directed the pilots to take actions that were incorrect based on the primary cockpit indications.

The actual malfunction presented in scenario #14 was an engine torque split, with the #1 engine failing to the low side. In this malfunction, the #1 engine was failing, and the #2 engine was providing power to keep the aircraft flying. The PADS incorrectly indicated that the malfunction was a #2 engine high-speed shaft failure. The procedure for resolving this malfunction includes an emergency shut down of the #2 engine, which in actuality is the only engine providing power. The information provided by the PADS for scenario #14 is shown in Figure 4.

- #2 Engine High Speed Shaft Failure
- COLLECTIVE ADJUST.
- EMERGENCY ENGINE SHUTDOWN ON #2 ENGINE.

Figure 4. PADS Display for Scenario 14.

The correct emergency procedure for the actual malfunction, an engine torque split, is shown in Figure 5, taken from the UH-60A Operator's Manual. In step 1, the emergency procedure requires the pilot to examine the Turbine Gas Temperature (TGT) indications to ascertain if one of the engines is overheating. If not, the pilot reduces power on the engine with the highest power indication (% TRQ). The pilot observes the results of this action and then must execute either step 2 or 3, depending on the reaction of the engines. If the % TRQ of the low power engine increases, this indicates that the high power engine is overspeeding and it is retarded (step 2). If the % TRQ of the low power engine does not increase, this indicates that the low power engine is failing, and the high power engine is returned to full power (step 3). Steps 1, 3, and 4 were the proper action for this malfunction.

1. If TGT limit on either engine is not exceeded, slowly retard Engine Power Control lever on high % TRQ engine and observe % TRQ of low power engine.

2. If % TRQ of low power engine increases, Engine Power Control lever on high power engine – Retard to maintain % TRQ approximately 10 % below other engine.

(OR)

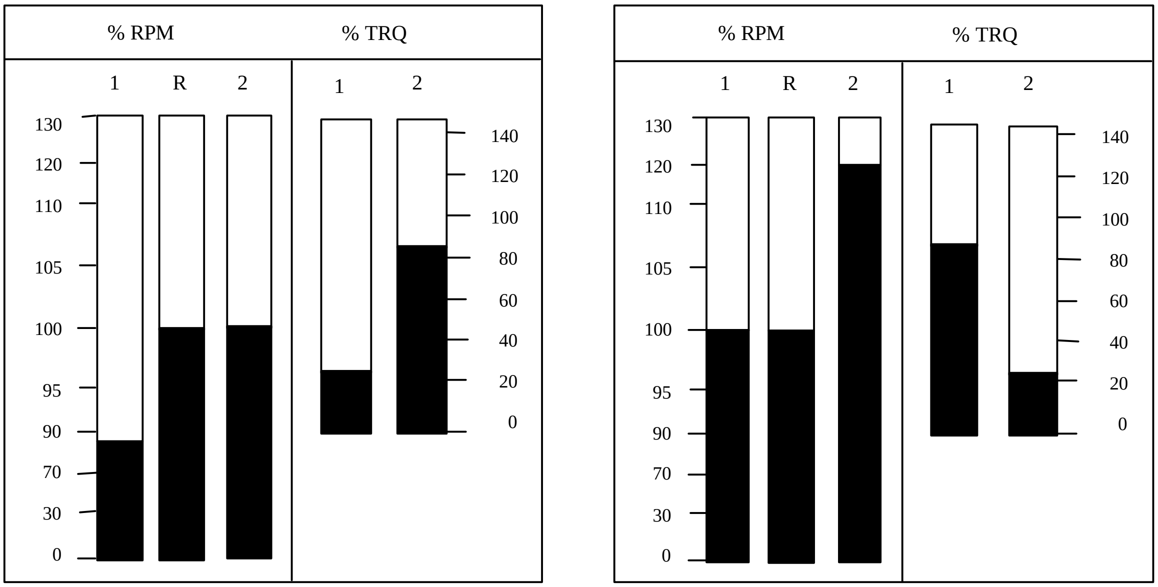
3. If % TRQ of low power engine does not increase, or % RPM R decreases, Engine Power Control lever – Return high power engine to FLY.

4. Land as soon as practicable

Figure 5. Emergency Procedure for Engine Torque Split

The two malfunctions have some similar indications and some contradicting indications. The #1 engine has a lower RPM indication in both malfunctions, so a first glance may verify the high-speed shaft failure malfunction indicated on the PADS display. However, the torque indications show the opposite indications of what would be indicated in a high-speed shaft failure. For this malfunction, the #2 torque indication would be low and #1 torque indication would be high. For the torque split (the actual malfunction), the #1 torque was low and the #2 was high, as shown in Figure 6. The pilots, therefore, were required to examine the instruments closely to ascertain that the malfunction presented by the PADS was not actually occurring, but that another failure was present.

If the subjects followed the PADS, and did not verify the malfunction with the instruments, they shut down the only good engine on the aircraft and crashed. If they attempted to verify the information on the display, they found that the instruments indicated an entirely different malfunction requiring a different procedure.



Cockpit indications for actual malfunction: engine torque split with #1 engine failing to the low side.

Cockpit indications for malfunction indicated on PADS: high-speed shaft failure on #2 engine.

Figure 6. Schematic of Cockpit Indications for Scenario #14

Experimental Procedure

Initial Briefings

The general scenario was described as an emergency mission to transport medical personnel to an aircraft crash site. The intent of this scenario was to add urgency to the mission to prevent the subjects from choosing to land and cancel the flight for minor malfunctions. The PADS display was explained as an experimental cockpit warning system. The tester reiterated to each subject the urgency of the mission and the functions and capabilities of the new warning system.

The subjects had three landing options during each scenario: the destination airfield one hour away, an alternate improved airfield 15 minutes away, or an unimproved emergency landing area (empty fields) in the immediate vicinity of the aircraft. The pilots were told that their mission was important, but passenger and aircraft safety were paramount.

The subjects were briefed that they were required to make all decisions in the cockpit. A confederate acted as second pilot. The second pilot acted as the pilot on the controls and took appropriate actions, but only at the direction of the subject pilot. Neither the experimenter nor the confederate discussed any failure situations or alerted the subject pilot to any problems.

Experiment Runs

Pilots were initially responsible for maintaining a heading and altitude to reach an airfield one hour away. They were presented with enroute system failure scenarios that required them to respond to the malfunctions and then make decisions on the status of the aircraft and its ability to continue the assigned mission. These scenarios were from both the primary and secondary experiments.

The pilots began each scenario in a straight and level flight mode enroute to their destination. System failures were initiated at random periods after the beginning of each scenario. When the simulator operator input the malfunction to the system, the aircraft systems and instruments reacted with the indicated failure. Simultaneous with the initiation of the malfunction in the simulator, the additional information was presented to the subjects on the LCD screen of the PADS in the cockpit. The malfunction information appeared almost simultaneously on the PADS display and on the aircraft instrumentation. The pilots were then to react to the malfunction by performing the appropriate emergency procedure. The malfunctions required accurate diagnosis and reaction in a short period of time to maintain the safety of the aircraft. At the conclusion of any immediate action, the pilots informed the copilot of their landing decision (continue mission, divert, or land immediately).

At the conclusion of each scenario, either after a landing was made, the decision was made to continue, or the aircraft crashed, the failure for that scenario was removed and the pilot either regained or was reset to his original flight parameters for the next scenario. This ensured that the starting conditions were the same for each scenario.

Debrief

After all 14 scenarios were complete, the pilots were asked to move to an adjoining room for a debrief. During the debrief, each scenario was discussed and a list of questions was answered. Paper copies of the information presented on the PADS display for each malfunction and the videotape of the simulator period were available to help the subjects recall specific scenarios when required.

Measures

<u>Performance Measures</u>

Three objective performance measures were selected for the experiment. These were:

- Did the subject take the correct action in response to the malfunction?
- Did the subject make the correct landing decision?
- How quickly did the subject respond?

Subjective Measures (Debriefing)

Each subject answered a series of questions in an extensive debrief following the simulation period. The debrief contained a list of questions for each scenario, followed by general questions on their opinions of the display and how they would improve it. These questions included:

- What was your first indication of a malfunction?
- What did you look at next to verify or gain more information?
- What was the primary indication you used to diagnose the system failure? (What indication- PADS, instruments, or other evidence, i.e., vibrations, control irregularities – did you use as the dominant factor in your diagnosis?)
- Was the new warning system helpful?
- Did the new warning system make your decision process faster (through additional information) or slower (due to additional time spent verifying)?

Results

The experiment resulted in 132 data points for the main experiment and twelve data points for the over-reliance experiment. One of the planned scenarios, #10-Airspeed Indications Incorrect, was not correctly presented by the simulator and was eliminated. Each subject therefore experienced eleven failures in the main experiment and two in the over-reliance experiment.

Main Experiment

Performance Measures

Due to the consistently high training level of the pilots, the performance measures from the primary experiment resulted in little measurable data. The subjects were over 92% correct in their responses to the malfunctions (122 of 132 correct), and the objective performance measures provided no measurable differences between test factors. The response time also provided little usable data. Determining when the response actually occurred was difficult because it could be manifested by several different actions (when the pilot vocalized the procedure he was taking, when he grabbed a switch, or when he completed the action) and therefore was judged to be too subjective to be usable.

Subjective Measures

The subjective measures examined were based on the subjects' responses to the questions asked during the debrief. The subjects were provided with multiple-choice responses for each question (i.e., PADS, instruments, or other), and the frequency of each response was recorded for each question. The frequency of each response was assumed to be an estimate of a binomial random variable. These frequencies were tested using standard t-tests to determine the significance of differences in the frequencies of the responses. A more conservative Tukey all-pairwise comparison test was performed on data with multiple factor levels. The results for the Tukey tests were identical with those from the individual t-tests.

Results Based on Knowledge Level. As the knowledge level of the system increased from the lowest to the highest level, the amount of information on the PADS display increased. At the higher levels of knowledge, the new warning system had more knowledge than was realistically possible with current technology. This may have affected the responses of the pilots. They may have been hesitant to rely on information that was beyond the known capabilities of the sensors on the aircraft (such as an impending failure, that could not be detected by current sensors).

The responses to the question, "What was your first indication of a malfunction?" are shown in Figure 7. The responses to the question, "What was the primary indication you used to diagnose the system failure?" are shown in Figure 8, where the "primary indication" was defined as the dominant piece of evidence that the subject used to make the diagnosis. The results of both the standard t-test and the Tukey test indicate that the knowledge level and the amount of information displayed by the system did not have any statistically significant affect on the use of the alerting system as the first indication or as the primary indication of a malfunction.

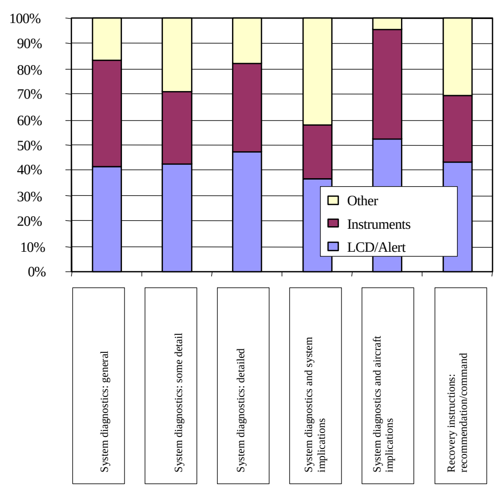


Figure 7. First Indication of Failure-by Knowledge Level

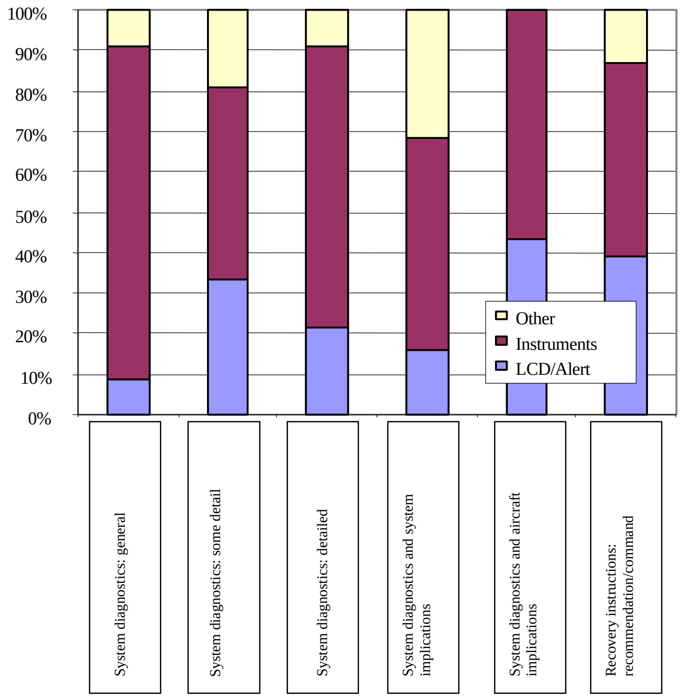


Figure 8. Primary Indication of Failure-by Knowledge Level

The responses to the question, "Was the new warning system helpful?" are shown in Figure 9. The percentage of responses indicating that the PADS was helpful increased steadily from the "System diagnostics: General" level through the "System diagnostics: Detailed" level. The responses then leveled off with no further improvement through the highest level ("Diagnostics, implications, and recovery instructions: Recommendation or command level"). In a paired-comparison statistical t-test at the 0.05 level of significance, the number of responses indicating that the system was helpful at the "System diagnostics: Detailed" level is significantly different than the number of responses indicating it was helpful at the "System diagnostics: General" level. The same test indicated that there was no difference between the "System diagnostics: Detailed" level through the three higher levels.

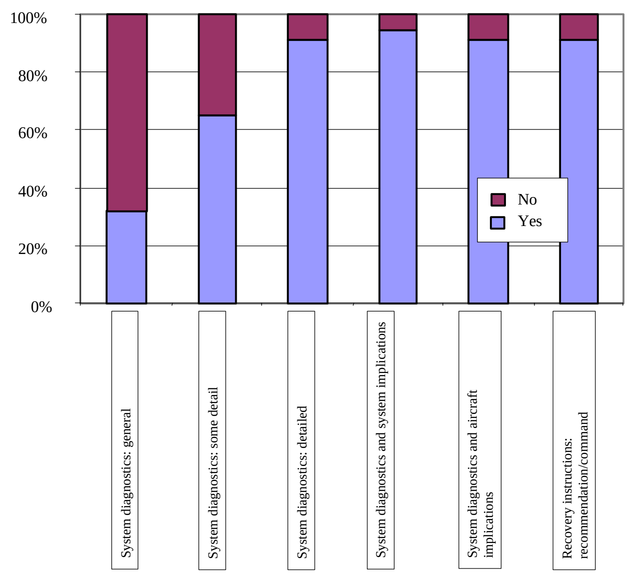


Figure 9. Alerting System Helpful-by Knowledge Level

The responses to the question, “Did the new warning system make your decision process faster (through additional information) or slower (due to additional time spent verifying)?” also revealed interesting results, as shown in Figure 10. The percentage of responses indicating that the PADS made the decision process faster did not increase as the knowledge level increased, as anticipated. Instead, the responses show no particular pattern through the “Detailed system diagnostics and aircraft implications” level. Then, comparing the “Detailed system diagnostics and aircraft implications” level and the “Diagnostics, implications, and recovery instructions: Recommendation or command” level, the responses for the highest level show a decrease in responses indicating that the PADS made the decision process faster, and an increase in responses that the PADS made the process slower. In a paired-comparison t-test at the 0.05 level of significance, the number of responses indicating that the PADS system made the decision process faster at the highest level (“Diagnostics, implications, and recovery instructions: Recommendation or command”) is significantly lower than the number of responses indicating it made the process faster at the next lower level (“Detailed system diagnostics and aircraft implications”). The responses indicating that the PADS system made the process faster are actually lower for the highest level (“Diagnostics, implications, and recovery instructions: Recommendation or command”) than for the “System diagnostics: Detailed” level. Finally, the percentage of responses indicating that the PADS system made the process slower is the highest of all levels at the highest level of knowledge (“Diagnostics, implications, and recovery instructions: Recommendation or command”).

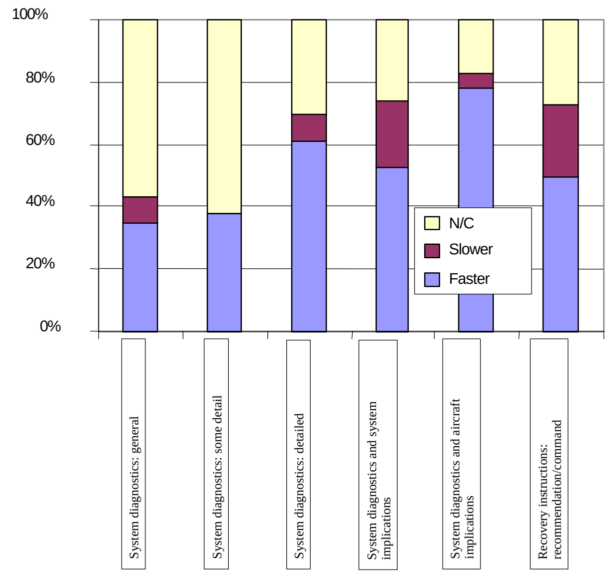


Figure 10. Decisions Faster or Slower-by Knowledge Level

Results Based on Assertiveness. Based on the assertiveness of the system, there was a significant difference in responses to the question, "What was your first indication of a malfunction?" In a paired-comparison t-test at the 0.05 level of significance, the number of responses indicating that the first indication of a malfunction was the PADS is significantly different for the subjects using the alerting system than for the subjects using the non-alerting system. These results are shown in Figure 11.

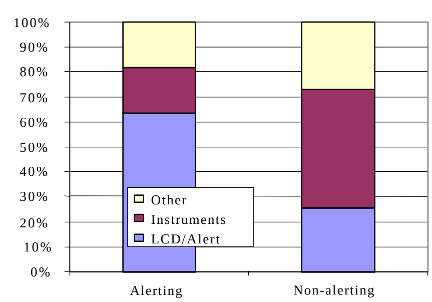


Figure 11. First Indication of Failure-by Assertiveness

In contrast, there was not a statistically significant difference in responses to the question, "What was the primary indication you used to diagnose the system failure?" In a paired-comparison t-test at the 0.05 level of significance, the number of responses indicating that the primary indication of a malfunction was the PADS is not significantly different for the subjects using the alerting system than for the number using the non-alerting system. These results are shown in Figure 12.

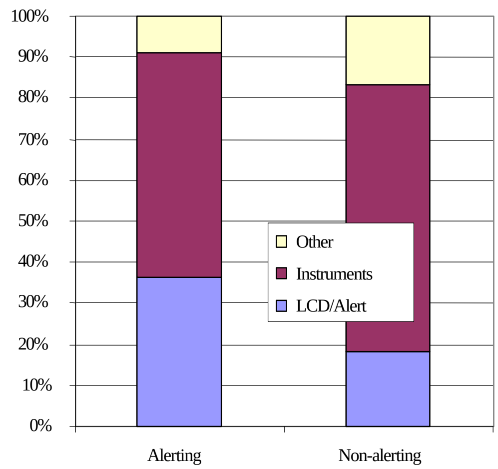


Figure 12. Primary Indication of Failure-by Assertiveness

<u>Results Based on Pilot Experience</u>. The results do not indicate that the experience of the pilots (measured in total hours of flight time or in hours of UH-60 flight time) had any affect on the use of the alerting system as the primary indication of a malfunction. The results were similar for all the subjective measures; pilot experience was not found to have a significant impact on any measure.

<u>Results Based on Order of Runs</u>. The results do not indicate that the order of the runs had any affect on the use of the alerting system as the first indication or the primary indication of a malfunction.

<u>Results Based on Scenario.</u> The results showed a significant difference in the responses based on the scenario being flown. In a paired-comparison statistical test at the 0.05 level of significance, the number of responses indicating that the primary indication of a malfunction was the PADS is significantly higher in the aggregate responses for the scenarios labeled as "less-trained" (7-12) than in the aggregate responses for the scenarios labeled "trained" (1-6). This may be a result of the homogeneity of the training of the subjects and the non-homogeneity of the level of training on the scenarios. The subjects generally responded similarly to the same malfunctions. For example, scenario 1, leak in the #1 hydraulic system, is a familiar malfunction that the pilots are

consistently trained on. The pilots were able to recognize and react to this malfunction almost immediately based on the cockpit warning system. There was a very low level of reported use of the PADS for this scenario. In contrast, scenario 11, crack in the tail rotor spar, is less familiar and not extensively trained on. It is also difficult to identify with cockpit indications. The reported use of the PADS for this scenario was significantly higher.

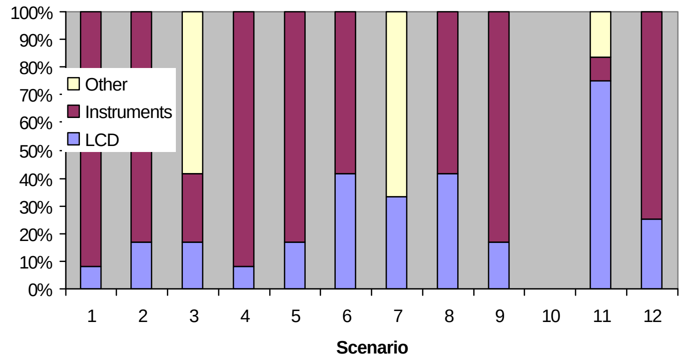


Figure 13. Primary Indication of Failure-by Scenario

Over-Reliance Test

All 12 subjects were presented with scenarios 13 and 14 at the end of their primary experiment runs. Scenario 13 was an impending main rotor failure with catastrophic results that could not be confirmed or disconfirmed with onboard instruments. All twelve subjects followed the recommendation of the PADS and landed immediately. Scenario 14 was an engine torque split malfunction in the simulator, with the PADS indicating a high-speed shaft failure, providing conflicting information. The results are shown in Table 2.

TABLE 2. Results of Scenario 14

Action	Number of Subjects
Disregarded LCD display	5
Followed LCD display	4
Initially followed LCD display	2
Landed with no action	1

Five subjects disregarded the PADS and reacted to the actual malfunction using the information on the cockpit instruments. These subjects performed the correct procedure for an engine torque split and continued the mission safely. Four subjects followed the PADS, shut down the good engine without confirmation from the other instruments, and crashed. Two subjects initially followed the PADS, decreased power or idled the good engine, then recognized that the instruments were not confirming the PADS information, and executed the correct procedure for a torque split. One subject initially followed the PADS and decreased power on the good engine, then stopped, and landed with no action. He stated that he "could not resolve the conflict."

Overall, seven of the twelve subjects performed the wrong procedure or were initially confused as to what action to take. The subjects had all experienced a similar malfunction earlier (torque split with the #2 engine failing), and all had performed the correct procedure. The subjects in each category were evenly divided by experience in total flight time.

Discussion and Conclusions

Although based to a large extent on subjective responses, the results of this study do suggest several conclusions. First, the results suggest that the subjects used the alerting system more as an attention-directing signal than as a diagnostic tool. There was a significant difference in the number of responses indicating that the PADS was used as the *first* indication of a malfunction for the subjects using the more-assertive alerting system than for the subjects using the non-alerting system. In contrast, the number of responses indicating that the PADS was the *primary* factor in deciding upon a diagnosis and resolution is not significantly different for the subjects using the alerting system than for the number using the non-alerting system. This may indicate that the higher assertiveness levels are useful for their alerting or attention-directing functions; however, it can not be assumed that the pilot will then follow the alerting system as the sole source of information. These results mirror those reported in studies of aviation operations, such as the delayed response problems with the early GPWS systems (DeCelles, 1991).

Second, the notion that "more information is always better" is disputed. In many domains a major benefit of information technology has been to provide the operator with more information, often integrated into a more useful form; in aviation the increased 'situation awareness' this can enable in the operator is considered a valuable contributor to safety. However, in this study the pilots were no more likely to use the PADS as the primary contributor to their decisions about diagnosis and resolution of hazard when the system provided high levels of knowledge than low levels of knowledge. Additionally, usefulness of the information provided by PADS was perceived by the subjects to reach a plateau at a limited amount of information. This was illustrated as the responses for the perceived helpfulness of the system increased significantly for the first three levels, then leveled off. This may indicate that the pilots utilized the additional increments of information up to a certain point, then stopped using it. Increasing amounts of information beyond that level were perceived as adding no additional benefit to the pilots.

Furthermore, additional information provided beyond that point may actually slow the decision making process. Responses indicated that increasing the level of information did not significantly improve the speed of the decision making process. In fact, the responses indicating that the PADS made the decision process faster were significantly lower at the highest level of knowledge (Diagnostics, implications, and recovery instructions: Recommendation or command) than at the next lower level (Detailed system diagnostics and aircraft implications).

These three results may be partially explained by the environment that PADS was tested in, including the substantial training and experience of these pilots. The UH-60, and its predecessors that these pilots are trained on, does not have sophisticated alerting systems. Information technology, such as the alerting systems examined in this paper, can provide several functions that at first glance appear to be potentially beneficial: more information; the integration of information; and decision aiding through the presentation of diagnoses and resolutions. However, the pilots are trained to react to malfunctions with specific immediate-action procedures through reference to a variety of cockpit indications, enabling the quick recognition-primed decision making process described in the introduction of this paper. Once the malfunction was identified, the pilots reacted with the prescribed responses. Beyond the initial alert to direct their attention to the problem, the additional information provided by PADS, such as why a system malfunctioned, the specifics of the malfunction, or the procedure to correct for the malfunction, may not have been necessary for a successful resolution to the problem. Instead, the additional information required a longer time to read and process, potentially placing an additional burden on the pilot in a high-tempo situation. A relationship between the usefulness of PADS and pilot training was noted in the post hoc analysis of pilot use of PADS between the different scenarios; a higher reliance on PADS was described by the pilots when faced with hazards with which they were not as familiar.

One possible implication of this finding is that the design advanced alerting systems should be guided by the training of the intended users, and could even potentially be used as surrogates for training. For complex systems with hundreds of possible malfunctions, the operators could be extensively trained on only the most common failures. The operators could depend on the alerting systems to diagnose and recommend solutions for the remainder of the failures. The system could be programmed to deliver only the information essential to the operator based on his training for each failure. The system would provide the minimal information necessary for the operators to identify the failures that they were trained to react to while providing more extensive information and recommended solutions for the uncommon and lesser trained problems. In each case, the recognition-primed decision making preferred by the pilots in this study could be followed: when faced with known failures, the operator can use cues from the environment to recognize the related resolution; and when faced with unknown failures, the alerting system can provide not only the cues, but also the answers.

On the other hand, no alerting system is correct 100% of the time. The over-reliance test in scenario 14 suggests that when an alerting system gives erroneous information that conflicts with other cockpit indications, serious mistakes can be made and pilots may have trouble correctly resolving the conflicts. Over 50% of the subjects had trouble determining the correct action when PADS presented an erroneous diagnosis in scenario 14. Several factors may have contributed to this problem. For example, the subjects may have quickly developed a level of trust in a new alerting system that allowed them to become complacent and unable to correctly react, or, in trying to find the cues to recognize the problem, they may have had difficulty in understanding an environment that was presenting conflicting information

This presents the designer of an alerting system with several dilemmas. While we would like the operator to recognize any hazard and respond correctly at all times, the dynamics of the situations in which they are making these decisions mandates a quick decision. Being both accurate and quick can be difficult for an operator; the most effective solution to date has been to require significant training of the operator so that he or she can react quickly and effectively. Whether used as a supplement to this decision process or as a training surrogate, alerting systems are not guaranteed to fundamentally make the operator take a more studied approach to diagnosis and resolution.

In addition, if the operators are no longer trained on recognizing and resolving the failures independent of the alerting system, they are no longer able to perform error-checking on the system and they may become entirely dependent on its functioning. This illustrates potential issues with the role of the alerting system. While alerting systems may be intended as aids, research on decision making has shown that salience has a large effect on what stimuli an operator will base his decisions on (Wickens, 1984; Mosier, *et al*, 1997). As more authoritative alerting systems are developed and provided a larger role, the tendency of the operators to make independent judgements and perform error-checking based on raw data will decline. As such, creating an alerting system whose commands can be easily assessed by the pilot remains a significant design challenge.

One approach may be to ameliorate the current perception that an alerting system is either 'right' or 'wrong'. For example, Hicks and Ross (1990) suggested attaching confidence levels to system output to increase pilot acceptance; likewise, Sorkin, *et al* (1988) recommended a likelihood alarm display to provide additional information to the pilot. The pilot can then merge his or her own evaluation with that of the system based on the amount of confidence the system projects in its output.

Another approach may be to have the system provide more explanatory information. Muir (1987) suggested that the first step in improving an operator's perception of trust in a decision aid is to train the operator to understand (to the extent possible in the application) how the decision aid works. Pritchett and Vandor (1999) demonstrated benefits from actively displaying to the operator the rationale behind automatic alerts.

The most significant consideration remains the intended role of the alerting system. Many different functions may be performed by the alerting system – attention directing, providing explanatory information, displaying diagnoses, commanding actions. Given the difficulty that an operator can have in resolving a hazard, and given time-critical reactions they require, care must be taken to ensure that the alerting system truly provides support and assistance to the operator, without burdening them with unwanted information and without placing them in the difficult position of assessing the correctness of alerting system commands.

Acknowledgements

This work was partly supported by the National Aeronautics and Space Administration / Langley Research Center under Grant NAG-1-2014. The authors would like to thank the following persons for their contributions to this study: Anna Trujillo and Paul Schutte at Langley Research

Center; the Flight Simulation Training Office at Fort Rucker; and the subjects who donated their time to participate in the experiment.

References

Deaton, J.E. & Glenn, F.A. (1999). The development of specifications for an automated monitoring system interface associated with aircraft condition. *The International Journal of Aviation Psychology*, 9(2), 175-187.

DeCelles, J.L. (1991). The delayed GPWS syndrome. Aviation Research and Education Foundation, Herndon, VA.

Endsley, Mica R. (1997) Level of automation: Integrating humans and automated systems. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*.

Hicks, M. & Ross, I. (1990). *Aircrew acceptance of automation in the cockpit.* AGARD CP-521 "Advance Aircraft Interfaces: The Machine Side of the Man-Machine Interface".

Klein, G. (1998). The recognition-primed decision model. In *Sources of power: How people make decisions* (pp.15-30), MIT Press.

Lee, J. & Moray, N. (1992) Trust, control strategies, and allocation of function in human-machine systems. *Ergonomics*. Vol 35 (10), 1243-1270.

Mosier, K.L. (1997). Myths of expert decision making and automated decision aids. In *Naturalistic Decision Making*, Lawrence Erlbaum Associates, Inc, Mahwah, NJ.

Mosier, K.L., Heers, S., Skitka, L.J., & Burdick, M. (1997). Patterns in the use of cockpit automation. In Mouloua, M. & Koonce, J.M. (Eds.), *Human-automation interaction: research and practice*, Hillsdale, NJ: Erlbaum.

Muir, B.M. (1987) Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27, 527-539.

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

Prevot, T. & Onken, R. (1996). In-flight evaluation of CASSY: A system providing intelligent on-board pilot assistance. *Air Traffic Control Quarterly*, Vol 3(3), 183-204.

Pritchett, A.R. and Vandor, B. (1999) Creating consonance between displays and alerting systems. Submitted to *Human Factors.*

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive Systems Engineering*. New York, NY: John Wiley and Sons.

Sorkin, R.D., Kantowitz, B.H., & Kantowitz, S.C. (1988). Likelihood alarm displays. *Human Factors*, 30(4), 445-459.

Wickens, C.D. (1984). *Engineering Psychology and Human Performance*. Columbus, OH: Charles Merrill.

---

* Mr. Davis is a Doctoral Sudent at the School of Industrial and Systems Engineering, Georgia Institute of Technology. He can be reached at: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA. Email: sdavis@isye.gatech.edu, Fax: (404) 894-2301.

**Dr. Pritchett is an Assistant Professor at the School of Industrial and Systems Engineering, Georgia Institute of Technology. She can be reached at: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA. Email: amy.pritchett@isye.gatech.edu, Phone: (404) 894-0199, Fax: (404) 894-2301.