

REACTIONS OF AIR TRANSPORT FLIGHT CREWS TO DISPLAYS OF WEATHER
DURING SIMULATED FLIGHT

FINAL REPORT

James P. Bliss, Ph.D.
Corey Fallon
Ernesto Bustamante
William R Bailey III
Brittany Anderson

Old Dominion University

National Aeronautics and Space Administration Grant No. NAG – 1 – 03061
(ODU Research Foundation Grant No. 130971)

Submitted as a Final Research Product to NASA Langley Research Center

January 31, 2005

Acknowledgements

The work reported within this document was supported by NASA Langley Research Center under Grant No. NAG - 1 - 03061 (ODU Research Foundation No. 130971). Grant Monitor was Dr. Jon Jonsson.

Any opinions findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

Abstract

Display of information in the cockpit has long been a challenge for aircraft designers. Given the limited space in which to present information, designers have had to be extremely selective about the types and amount of flight related information to present to pilots. The general goal of cockpit display design and implementation is to ensure that displays present information that is timely, useful, and helpful. This suggests that displays should facilitate the management of perceived workload, and should allow maximal situation awareness. The formatting of current and projected weather displays represents a unique challenge. As technologies have been developed to increase the variety and capabilities of weather information available to flight crews, factors such as conflicting weather representations and increased decision importance have increased the likelihood for errors. However, if formatted optimally, it is possible that next generation weather displays could allow for clearer indications of weather trends such as developing or decaying weather patterns. Important issues to address include the integration of weather information sources, flight crew trust of displayed weather information, and the teamed reactivity of flight crews to displays of weather. Past studies of weather display reactivity and formatting have not adequately addressed these issues; in part because experimental stimuli have not approximated the complexity of modern weather displays, and in part because they have not used realistic experimental tasks or participants. The goal of the research reported here was to investigate the influence of onboard and NEXRAD agreement, range to the simulated potential weather event, and the pilot flying on flight crew deviation decisions, perceived workload, and perceived situation awareness. Fifteen pilot-copilot teams were required to fly a simulated route while reacting to weather events presented in two graphical formats on a separate visual display. Measures of flight crew reactions included performance-based measures such as deviation decision accuracy, and judgment-based measures such as perceived decision confidence, workload, situation awareness, and display trust. Results demonstrated that pilots adopted a conservative reaction strategy, often choosing to deviate from weather rather than ride through it. When onboard and NEXRAD displays did not agree, flight crews reacted in a complex manner, trusting the onboard system more but using the NEXRAD system to augment their situation awareness. Distance to weather reduced situation awareness and heightened workload levels. Overall, flight crews tended to adopt a participative leadership style marked by open communication. These results suggest that future weather displays should exploit the existing benefits of NEXRAD presentation for situation awareness while retaining the display structure and logic inherent in the onboard system.

REACTIONS OF AIR TRANSPORT FLIGHT CREWS TO DISPLAYS OF WEATHER DURING SIMULATED FLIGHT

INTRODUCTION

For many years, researchers and designers have studied strategies for incorporating information displays within the aviation cockpit. As noted by Stokes and Wickens (1988), displays are often the primary way that pilots are made aware of events happening outside of the cockpit. However, for situational information to be effectively transferred to the pilot(s), designers must overcome a host of challenges associated with incorporating displays in airplane cockpits.

The general goal of cockpit display design and implementation is to ensure that displays present information that is timely, useful, and helpful. Because pilots frequently exercise divided and selective attention during flight, displays must never constitute a cognitive burden on the pilot. Researchers have focused on two criteria to ensure that displays are cognitively manageable. The first is mental workload. The concept of workload has been discussed for a number of years (see Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, & Wherry, 1989); however, its precise definition is still elusive. Hart and Wickens (1990) proposed a definition that refers to workload as the mental cost of accomplishing a task, where additional workload reduces an operator's ability to accomplish a task, all else being equal. Kantowitz & Casper (1988) discuss the fact that automation, including automated displays, may decrease workload because certain tasks are relegated to automatic control and monitoring. However, automation may also increase workload because cognitive monitoring must be increased to keep track of the system.

Another criterion by which displays are evaluated is their ability to foster full situation awareness by the pilot. In one of the first treatments of the concept of situation awareness, Endsley (1990) characterized it as including a consideration of present and future events germane to an ongoing task. As discussed by Tsang and Vidulich (2003), perhaps the most critical driver of situation awareness for pilots is the quality of information obtained about the flight environment. Because of this, there is a natural and direct relationship between the effectiveness of displays and the resulting level of situation awareness.

Ultimately, the product of flight displays is a decision or series of decisions to be made by the flight crew. Making such decisions has always been difficult, particularly in recent years as the skies have become more congested and flight capabilities have increased. Pilots must make decisions about flying, navigating, and communicating based on a number of variables. However, as Reason's (1990) influential decision making model suggests, the decisions are often fraught with uncertainty, particularly when the flight conditions do not mesh well with established or trained scenarios.

Primary among information display challenges is the idea that displayed parameters should be interpretable by the flight crew. Therefore, much research has been devoted to identifying principles for formatting information on displays. As the result of several decades of work by numerous researchers, Wickens (2003) proposes seven principles to drive display design:

- Information Need - Pilots should be presented with only the most critical information necessary to complete a task. Excessive information should be avoided.
- Legibility – While display component legibility does not guarantee usefulness, it is necessary that pilots be able to make sense of the displayed components.
- Proximity Compatibility – Designers should put sources of information requiring integration by the pilot close together, and should put the most critical sources of information directly in front of the pilot.
- Pictorial Realism – Displays should accurately reflect their real-world analogs.
- Principle of the Moving Part – Related to the Principle of Pictorial Realism, this principle states that moving display components should move in a manner similar to the moving element in the pilot's mental model.
- Predictive Aiding - Displays that offer information about projected future events and states should be as accurate as possible.
- Discriminability – Displays should be distinct from each other, particularly if they might be included in the same context.

Cockpit Displays of Weather

Some of the biggest challenges concerning cockpit displays involve weather information presentation. Commercial flight crews must make a number of important decisions when interpreting weather information. The quality of these decisions could potentially impact flight safety, passenger comfort, fuel consumption and flight time. However, as technologies have been developed to increase the variety and capabilities of weather information available to flight crews, factors such as conflicting weather representations and increased decision importance have increased the likelihood for errors. Some researchers have suggested that increased communication and effective leadership styles may be effective in reducing errors under such conditions (Foushee, 1982), and have recommended that empirical research be conducted to examine the effects of communication and leadership styles on flight crews' reactions to advanced weather displays.

Because of the large number of available weather information sources, and the varying reliability, validity, urgency, and relevance of those sources, pilots are required to make frequent judgments under conditions that may be less than ideal. For that reason, some researchers have created tools to improve the pilot decision making process. The Weatherwise program (Wiggins, 1999) is an example of such a tool. This program allows pilots to practice interpreting weather related information and formulate decisions based on that information.

Historically, pilots have obtained weather related information from a large variety of sources. They consult TAFs, METARs, FAs, AIRMETs, NOTAMs, PIREPs, SIGMETs, Onboard and NEXRAD radar information, and information about winds aloft, icing, lightning, convective activity, hail, and precipitation. Based on examinations of these sources of information, pilots must sometimes make weather-induced route deviation decisions. Fortunately, recent technology has allowed weather information to become widely available in a graphical format. An example of this is the availability of weather information from ground-based radar sources, such as Next Generation Radar (NEXRAD), in the cockpit. NEXRAD is a Doppler radar that integrates information

about wind and precipitation on a single graphical display. This technology allows weather information to be updated often, so that pilots can determine past, present, and future weather states, and plan accordingly.

Integrated graphical weather displays, such as NEXRAD, offer a number of potential advantages. Many researchers have suggested that presenting multiple weather information sources concurrently may allow flight crews to draw conclusions with minimal cognitive expenditure. For example, the Proximity Compatibility Principle suggests that displaying multiple parameters may constitute less of a processing drain on aircrew members (O'Brien & Wickens, 1997). In addition, Wickens (2000) notes that integrated displays may reduce drains on selective attention because multiple sources of weather information that normally compete for flight crews' attention may be viewed concurrently.

In addition to consolidating multiple sources of weather information, integrated graphical weather displays are also capable of presenting updated weather information in near-real-time. Sherman (2003) suggests that the reliability of weather information might be improved by presenting it rapidly to the flight crews. In turn, this may also improve flight crews' interpretations of weather trends such as developing or decaying weather patterns (Boyer, Campbell, May, Merwin, & Wickens, 1995). In general, integrated graphical weather displays may allow pilots to manage risk more effectively (Orasanu, Davison, & Fischer, 2001). In fact, O'Brien and Wickens (1997) have demonstrated that in some circumstances pilots can make better judgments from integrated information than when such information is presented separately.

Despite the aforementioned benefits of using integrated graphical weather displays, problems may arise when they are combined with other sources of weather information, such as the aircraft's onboard radar. Lindholm (1999) notes that combining integrated graphical weather information with other displays may result in excessive visual clutter that can cause information to be undetected or misinterpreted over time. Wickens and his colleagues have also noted the implications of excessive visual clutter on performance indices such as workload, errors, and response time (Wickens, Kroft, & Yeh, 2000).

Because of the large number of available weather information sources, and the varying reliability, validity, urgency, and relevance of those sources, pilots are required to make frequent judgments under conditions that may be less than ideal. According to Sly & Harmann (1999), weather induced route deviation decisions are influenced by a number of variables, including the following:

- **Type of Hazard** – Wind, precipitation, lightning, and turbulence may have differing effects on aircraft aerodynamics and flight parameters. Therefore, each weather element may hold varying implications for the current flight plan.
- **Distance or Time in Weather** – As with most threats to flying safety, the longer or farther that the aircraft flies under hazardous conditions, the greater the overall danger level.
- **Probability of Hazard Occurrence** – Considering the development of hazards is especially important because of the strategic nature of flight planning.

- Coverage or Density of Hazard – Pilots need to know what types of deviation maneuvers they can make, so it is important to understand the extensiveness of the weather related problem.
- Personal Preferences – Based on personal risk taking, experience, and prior knowledge, pilots often vary with regard to their willingness to fly into or around weather hazards.
- Fleet Wide Optimization – In many cases, air transport carriers are encouraged to conserve fuel and to reach their destinations in an expeditious manner. These concerns will likely impact the choices made by pilots.
- Mission Constraints – Sometimes, a route deviation will interfere with air space restrictions, or will require extra time to complete. Similarly, the number of passengers may influence a pilot's decision to enter a weather area.
- Carrier Philosophy – Air transport companies often have their own philosophies that influence whether pilots should fly through weather areas.
- Aircraft Type – In most cases, larger aircraft are less affected by weather variables than small aircraft.
- Severity of Weather – This basic consideration is generally considered first, and generally interacts with several of the other variables listed above.

Compounding the complexity of weather induced route deviation decisions are the variety and capabilities of weather information available for consultation by the pilot. Not only do pilots have access to weather information in a large number of formats, available technology may actually provide an overabundance of detailed information. For example, NEXRAD is capable of measuring wind information out to 60 nm and other weather features out to 130 nm. Although this increases the potential to receive detailed weather forecasts in the cockpit, such excessive detail may be misleading because of the rapidly changing nature of weather. As technologies for weather display continue to mature, choices may have to be made to avoid increasing mental workload while optimizing situation awareness and decision quality. Specifically, designers of weather display systems may opt to let pilots choose which of several types of weather information sources are available at any given moment.

Sly and Hartmann (1999) conducted a number of interviews with aviation weather experts and airline representatives to determine the weather related factors that pilots consider most important for route deviation decisions. These factors were subsequently the focus of an aviation weather information decision aid created by Honeywell, Inc. The results of those interviews suggested that pilots paid most attention to convection, icing, turbulence, volcanic ash, and ozone concentration levels, in that order. Sly et al.'s (1999) work is meant to facilitate decisions about weather information consolidation. As demonstrated by early research concerning object displays (Jacob, Egeth, & Bevon, 1976), it is possible to create information displays that include a number of variables so that global principles can be comprehensible to task operators.

As suggested earlier, a central issue surrounding the use of weather displays in the cockpit has been the integration of various types of weather information, and the integration of weather displays with other existing displays in the cockpit. Wickens (2000) has discussed the benefits of integration with other displays, noting that such

integration may reduce drains on selective attention, because competitors for attention (e.g., weather and traffic) can be viewed concurrently.

Certainly, integration is an important issue, because implementation of weather displays must occur within the context of an extremely crowded cockpit. Weather displays will have to compete with the primary flight display, attitude indicators, altitude gauges, and other existing status and warning displays for the pilot's attention. Wickens and his colleagues have provided a theoretical foundation for studying these issues. However, more empirical research must be completed for designers to have a set of guidelines for display construction and implementation.

There are many potential advantages to displaying multiple sources of weather data concurrently. If formatted optimally, it is possible that combined displays could allow for clearer indications of weather trends such as developing or decaying weather patterns (Boyer, Campbell, May, Merwin, & Wickens, 1995). It is also possible that more weather information might be presented rapidly to the cockpit, thereby avoiding the problems discussed by Sherman (2003) related to the lack of reliable weather information. There may well be a cognitive processing advantage associated with weather information consolidation as well. The Proximity Compatibility Principle suggests that displays including multiple parameters may constitute less of a processing drain on aircrew members (O'Brien & Wickens, 1997). Similarly, presenting a number of weather variables to pilots concurrently may allow for conclusions to be drawn with minimal cognitive expenditure. Another potential advantage is particularly relevant for non-instrument rated pilots. Incursion from visual flight rules conditions to instrument meteorological conditions (VFR to IMC) is a frequent factor implicated in general aviation accidents. As noted by Goh and Wiegmann (2001), one reason that pilots make VFR to IMC incursions is because they overestimate visibility in adverse weather. Combining weather information sources may allow for more realistic judgments to be made, thereby reducing the number of VFR to IMC incursions. In general, combining weather information on cockpit displays may allow pilots to manage risk more effectively (Orasanu, Davison, & Fischer, 2001).

Although advantages exist for combining weather related information in the cockpit, there are also potential disadvantages. As noted by Wickens and his colleagues (O'Brien & Wickens, 1997), graphical depictions of weather phenomena should be constructed and implemented with care, lest they fail to portray aspects of weather accurately. A pertinent example of such failures is certain weather software programs that fail to show cloud top altitude information. Because pilots may elect to fly above weather, it is crucial to understand the upper boundaries of weather. Another disadvantage concerns the varying rates at which component weather information might be updated. Some weather information may be updated as frequently as every minute or two. In contrast, other weather related information, particularly if obtained using a "request/reply" datalink system, may be updated relatively infrequently (<http://www.avidyne.com/narrowcast/Narrowcast.htm>). Pilots may incorrectly assume that all weather components are updated with the same frequency, and subsequently make incorrect decisions based on that erroneous information.

A particularly troublesome by-product of combining weather data sources is graphical clutter. Not only do weather displays themselves often resemble a jumbled mess of information, Lindholm (1999) discusses the need to integrate weather

information with existing cockpit displays, a possibility that has been explored by other researchers and developers (see O'Brien et al., 1997). Lindholm (1999) notes that integrating weather display information with other displays such as Cockpit Displays of Traffic Information (CDTI) and Enhanced Ground Proximity Warning Systems (EGPWS) may result in unacceptable levels of visual clutter. Such clutter would likely be incongruent with effective cognitive processing, and may violate the proximity compatibility principle (Wickens & Hollands, 2000).

Trust of Displayed Weather Information

There are other more general problems with integrated displays of weather information as well. The onboard and NEXRAD radar systems differ with respect to their degree of complexity and capabilities. These differences may produce conflicting weather representations between the two systems. Such conflicts may decrease the reliability of weather information made available to flight crews. This in turn can lead to overtrust or undertrust in either of those systems. Related to this, one of the chief concerns is whether operators (pilots) will trust the alarm signals that are typically associated with weather information displays. Muir (1994) as well as Lee and Moray (1994), have suggested that effective human interaction with technology must include a degree of trust. However, individuals have been shown to overtrust or undertrust automated systems, and in turn exhibit degraded performance (Parasuraman & Riley, 1997).

To help pilots make sense of the continuous data that are presented to them, designers often incorporate visual and/or auditory alarm signals to indicate when weather phenomena are sufficiently critical to warrant special attention. However, Bliss and his colleagues (Bliss, 1993) have demonstrated that operators may mistrust alarm systems that demonstrate frequent false alarms. That mistrust is then manifested in degraded task performance speed and accuracy. In the case of automated weather display systems, pilots may not trust the information given because of variations in data age, comprehensiveness, urgency, or redundancy. Although researchers have studied some of these variables in generic laboratory settings (see Bliss, Deaton, & Gilson, 1995; Bliss, Jeans, & Prioux, 1996) replication with weather displays and alarm signals is needed.

Display Reactions by Teams

Another area that has been overlooked by researchers until recently concerns the responsiveness of teams to displayed weather information. Almost without exception, the principles of display design discussed earlier were formulated from research using individual participants. In very few cases have researchers investigated how teams (for example, a pilot-copilot team) might react to variations in weather display format. Given the work by Foushee (1982) and others regarding cockpit resource management (CRM), it is likely that respondents in teams would display marked tendencies dependent upon leadership, communication, cohesiveness, and a variety of other factors. More empirical work is needed to isolate and quantify the effects of such social factors on weather display interpretation and responsiveness.

Many psychologists today believe that training focused on team interdependence and communication is the best way to prevent disorder and maintain the lines of communication (Sexton & Helmreich, 2000; Wickens 1995). For example, Foushee's

(1982) research regarding cockpit resource management has suggested that flight crews' responsiveness to weather displays is likely to depend on the leadership styles and communication tendencies of the flight crew. Thus, it is likely that the pilots' conceptualizations of display urgency and reliability may be mediated by the addition of another crewmember. Bliss (2003) has demonstrated this possibility for alarm signal reliability.

Even though Foushee (1982) and others have pointed out the desirability of effective communication and coordination between members of the flight crew, researchers have typically assessed only individual reactions to cockpit displays. Given the importance of teamed decision making in the cockpit (Foushee, 1982), it is important to investigate the impact of integrated graphical weather displays on teamed decision making, particularly when the weather information is not completely reliable. Thus, more empirical work is needed to isolate and quantify the effects of social factors such as communication on weather display interpretation and responsiveness.

However, communication among flight crew members can vary widely in style and its effects on performance. Effective communication in the cockpit is often determined by the leadership style of the pilot. Therefore, by studying various pilot leadership styles, it may be possible to determine which type of communication works best during situations of disorder and ambiguity, where the potential for human error is greatest. Thus, pilot leadership style may be an essential component in the management of human error (Helmreich, Merritt, & Wilhelm, 1999).

Normative Decision Theory

In the early 1970s Vroom and Yetton developed a theory that characterizes two styles of leadership, participative and autocratic, (Chemers, 2000). According to Normative Decision Theory, the participative leader promotes two-way communication and allows others to have equal influence in the decision making process. Conversely, the autocratic leadership style involves very little communication between the leader and other team members. Therefore, an autocratic leader makes all of the team decisions without much input from other team members (Chemers, 2000).

Normative Decision Theory states that the effectiveness of each style of leadership is dependent on the situation (Chemers, 2000). According to this theory, the autocratic style is more effective in situations where the tasks are clear, and optimal choices are obvious. In these situations very little communication is needed in the decision making process, allowing the autocratic leader to make quick decisions (Vroom, 2000).

The participative style is more effective in an ambiguous environment when the optimal decision is not readily apparent. This style is also more effective when faced with very important decisions (Vroom, 2000). The increased level of two-way communication may help to clarify the situation and improve teamed decision making (Chemers, 2000). Recent studies on leadership in the cockpit have indicated that participative leadership may be effective at minimizing the number of errors made in the cockpit (Foushee, 1982, 1984; Sexton & Helmreich, 2000). In addition, Nicholas and Penwell (1995) examined the leadership styles of aviators and found that the more effective leaders employed a predominantly participative leadership style and a strict autocratic style "does not lend

itself to effective operation of complex, technical machinery" (Nicholas & Penwell, 1995, p.70).

Past Weather Display Research

In the years since technology has enabled the design of integrated, graphical weather displays, there have been relatively few studies of their effectiveness. Those research studies that have been completed have typically concerned proper formatting of weather information and the impact of graphical weather displays on pilot decisions.

Wickens and his colleagues have generally led the way, investigating display structure, display element compatibility, and implications for pilot attention. As described in Wickens' chapter within Tsang and Vidulich's recent text (2003), a central issue has been the existence of clutter on weather displays (p. 164). Wickens and his colleagues have noted the implications of excessive display clutter on such performance indices as workload, time delay, and errors (Wickens, Kroft, & Yeh, 2000). Another focus has been determining optimal strategies for organizing complex displays so that pilots can navigate through them without becoming cognitively lost (Roske-Hofstrand & Paap, 1986). One insidious aspect of clutter actually concerns the decluttering process. As Wickens (2003) notes, the act of removing clutter from a display, as might be done by a computerized filtering system, may actually cause a pilot to fail to notice important information.

Because weather is just as important to general aviation pilots (perhaps moreso) as transport pilots, some researchers have studied weather influences on GA accidents. Capobianco and Lee (2001), for example, conducted an archival analysis of the Aviation Safety Reporting System. In their work, they searched from 1995 to 1998, isolating 1528 accidents where the phrase "weather condition" was used in the narrative. After performing an in-depth analysis of the narrative contents, the authors concluded that VFR to IMC incursion plays a major role in aviation accidents. Particularly troubling for GA pilots were weather conditions including low cloud ceilings, fog, high and variable wind conditions, and flight during darkness.

Other researchers have addressed weather display issues as well. For example, Beringer and Schvaneveldt (2002) recently described research conducted at the Civil Aerospace Medical Institute to determine the weather information required by pilots during various stages of flight. Such information is important to know, as designers strive to meet the weather display needs of pilots. It may influence how and when weather information is displayed. However, as noted already, the integration of these sources of information is still a troubling issue for designers.

Other recent research, reported by Latorella and Chamberlain (2002) has concerned the perceived risk associated with a variety of weather events. Such information may interact significantly with weather display reliability. Pilots may become overly conservative if they perceive that unreliable weather displays signal risky events. Conversely, pilot attitudes toward risk may be mediated by unwarranted trust in display systems (Parasuraman & Riley, 1997).

A consideration of available research literature shows that research is needed to clarify the role of variables on display of weather information in the automated cockpit. Since the early 1990s, scientists at NASA's Langley Research Center have been committed to answering the questions that exist concerning weather display

implementation and interpretation (Scanlon, 1992). Funded by NASA's AWIN program, numerous researchers have begun to investigate various issues concerning weather display. Langley's Crew/Vehicle Integration Branch Level 3 Research Plan makes clear the necessity for behavioral research to assess the impact of new and advanced display formats on pilot preferences and performance in the cockpit. Included within that document is a series of planned evaluations to determine the effectiveness of various elements of weather displays. Evaluations are planned for systems to be implemented in transport, commercial, and general aviation environments. Initially, a flight test was planned for 2005, during which selected research findings would be replicated and evaluated. Recommendations would then be made to the Federal Aviation Administration (FAA) and other customers regarding the acceptability of particular weather display configurations and general implementation strategies.

NASA enlisted the aid of several organizations to meet their research goals. As the ultimate customer, the FAA is very interested in the process and outcomes of NASA's program of behavioral research. In many cases, they broadly define the research processes to be followed, and the deliverables resulting from those processes. The National Center for Atmospheric Research (NCAR) designs sensor/alerting algorithms to control the onset and offset of alarm signals. Industrial partners such as Rockwell Collins, Honeywell, and PPI Aviation construct prototype display concepts, and provide those to NASA for testing and evaluation. Academic and simulation enterprises such as Wichita State University, Georgia Tech Research Institute (GTRI) and Research Triangle Institute (RTI) collaborate with NASA to conduct simulations and evaluations of the designed concepts. The common interest that these organizations share is a desire to optimize the interface between weather technology and the human users of that technology.

On July 10-11, 2002, the third meeting of the FAA/NASA Human Factors Weather Research Coordination Effort was held at Langley Research Center. In attendance were representatives from NASA, the FAA, Academia, and several industries involved in the design and implementation of weather displays in the cockpit.

During that meeting, representatives discussed the status and progress of several funded research projects, and highlighted program management and research issues that warranted continued examination. Those issues included the following:

Formatting of display elements in the cockpit – There was widespread agreement among the meeting participants that display factors such as colors, icons, symbology, and text currently lack standardization. For weather displays to be useful by pilots, it is critical that such standardization be implemented, and that the strategies used to construct textual and graphical displays be based on sound human factors research.

The impact of advanced weather displays on flight crew workload – As noted by Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, and Wherry (1989), operator workload is a critical issue. This is particularly the case in aviation. Flight crews are responsible for a tremendous number of tasks, particularly during aircraft takeoff and landing. It is important that designers of weather displays consider their impact on workload, and that they attempt to measure that impact as precisely as possible before implementing displays.

The impact of advanced weather displays on flight crew situation awareness – The importance of situation awareness for pilots has been recognized since the early 1990s. In 1995, a conference was held in Orlando, Florida, leading to the publication of a special issue of *Human Factors* devoted to the topic of situation awareness. For pilots, an important component of situation awareness is cognizance about the presence and magnitude of severe weather along the flight path. Those who design and implement weather displays should attempt to maximize situation awareness.

Alerting algorithms and stimuli within advanced weather displays – The role of weather displays is to present near-real-time (“nowcasting”) and predictive information about weather anomalies. In many cases, such presentation includes generating visual and auditory alarm signals to draw the flight crew’s attention to potential weather-related problems. A variety of issues surround the effective implementation of alarm stimuli in the cockpit. Researchers have discussed the importance of proper urgency formatting (Edworthy, Loxley & Dennis, 1991), presence of collateral alarm signals (McDonald, Gilson, & Mouloua, 1996; Bliss & Capobianco, 2003) and signal reliability (Bliss, Gilson, & Deaton, 1995; Getty, Swets, Pickett, & Gonthier, 1995), to name a few. Of these, reliability is a particularly complex issue, because it is dependent upon a host of factors, including age and comprehensiveness of weather information.

Collaborative decision making in reaction to weather information – Even though Foushee (1982) and others have pointed out the desirability of effective communication and coordination between members of the flight crew, researchers typically assess only individual reactions to cockpit displays. It is likely that the influence of workload and situation awareness on pilot performance, as well as pilot conceptualizations of display urgency and reliability, may be mediated by the addition of another crew member.

As mentioned in the HF WX Workshop, there are a number of information sources that pilots rely on to learn about weather conditions. Researchers have determined that pilots use some of these sources of information (e.g., PIREPS) more often than others, and that some weather elements such as lightning are not as threatening to flying aircraft as others, such as hail. To work within the constraints of cockpit display space, designers must often choose among these weather sources, determining what are the most critical weather factors to show pilots during flight. In many cases, displays are developed that present a number of weather factors at the same time, or on alternate screens. Such displays must be evaluated to ensure that useful information is conveyed in an intuitive and memorable manner.

Goal of this Research

It is clear that there are a host of issues to be investigated before multifaceted displays of weather information in the cockpit may be considered successful. The research project described within this report was undertaken to examine several of these issues. Given the importance of teamed decision making in the cockpit (Foushee, 1982) and the likely impact of weather displays on teamed decision making, it is important to investigate issues related to such teamed weather decision making, particularly when weather information is not completely reliable or current. It is also important to

investigate the levels of trust that flight crews might exhibit toward existing and planned displays of weather information.

In this research, pilot/copilot flight crew teams were required to fly a simulated route, encountering mock weather events while in flight. At 160, 80, 40, and 20 miles from a simulated weather event, crews were presented with a combination of onboard and NEXRAD weather imagery to facilitate their decision making. At times, these sources of weather information agreed with each other. At other times, they did not. In reaction to the displayed weather events, crews were required to discuss and decide whether and how to make course deviations. They also provided data concerning their trust of the displayed information sources, their workload, and their situation awareness whenever the weather events were displayed. At some times the seating arrangement dictated that the captain was the pilot flying; at other times the first officer was the pilot flying.

Hypotheses

The main dependent measures we collected each time weather was displayed may be conceptualized as qualitative judgment data or quantitative performance data. Specifically, deviation decision accuracy (correctness of deviation decisions) was a quantitative performance variable, whereas pilot confidence, perceived situation awareness, perceived workload, and perceived display trust were qualitative judgment variables.

Regarding deviation decision accuracy, we expected that decision accuracy would be greater when the onboard and NEXRAD sources of weather were in agreement than when they were not. This hypothesis would be consistent with available theories of machine trust (Muir, 1989) suggesting that redundant displays of information are trusted and reacted to more readily (Selcon, Taylor, & Shadrake, 1991; Bliss, Jeans, & Prioux, 1996). We made no hypotheses concerning whether accuracy would improve as a function of range to weather or pilot flying, because of the possibility for other factors to mediate these relationships.

Concerning the qualitative judgment data for display trust, workload and situation awareness, we expected flight crews to show more trust, lower workload and greater situation awareness when displays agreed with each other. However, we made no prediction concerning these measures as a function of range to weather or pilot flying, due to the expected complexity of interactional influences.

In the context of the experiment, we paid particular attention to the level and type of communication that occurred within flight crew participants. With regard to our assessment of communication patterns among the flight crews, we made several hypotheses. Our first hypothesis was that conflicting information presented on the onboard radar and NEXRAD displays would be associated with the participatory leadership style as distance from the weather event was reduced. We expected this hypothesis to be supported if the flight crews displayed the participative leadership style characteristics described by the Normative Decision Theory. Vroom (2000) suggests that participative leadership works best in ambiguous work environments, or when teams are faced with important decisions. For this study, presenting conflicting weather information generated ambiguity, and decision importance increased as the flight crews approached the weather event. An analysis of variance could not be used to test this

hypothesis because leadership style was a dichotomous dependent variable. Therefore, a logistic regression analysis was used to test the hypothesis.

The second hypothesis was that conflicting information presented on the onboard radar and NEXRAD displays would be associated with high levels of communication as distance from the weather event was reduced. This hypothesis would have been supported if flight crews displayed the participative leadership style tendencies. Chemers (2000) suggests that participative leaders promote more two-way interactions than authoritative leaders. Therefore, in conditions where flight crews encountered ambiguous information (i.e., conflicting weather information) and were faced with an important decision (i.e. closer to the weather event), higher levels of communication should have been present. As with the first hypothesis, an analysis of variance was not appropriate for testing this hypothesis, because the dependent variable was ordinal. Therefore, an ordinal logistic regression was computed.

The third hypothesis was that conflicting information presented by the onboard radar and NEXRAD displays would not be associated with more decision errors as distance from the weather event is reduced. This hypothesis was to be tested only if one or both of the previous hypotheses was supported. Vroom and Jago (1978) suggested that decision errors would be less frequent when the principles of the Normative Decision Theory are correctly applied. Therefore, if participatory leadership or increased levels of communication were present during ambiguous situations where teams are faced with an important decision, then decision errors were not expected to be significantly higher under those conditions. This hypothesis also includes a dichotomous dependent variable, so a logistic regression was run.

METHOD

To determine the influence of the independent variables described earlier, the researchers used a 4x3x2 within-groups design. Distance to weather was a within-groups independent variable with four levels: weather display at 160, 80, 40 and 20 nautical miles away from the potential weather event. A second independent variable was the agreement between the Onboard and NEXRAD sources of weather information. The agreement independent variable had three levels was manipulated within groups. The three levels of this variable included situations when only the onboard radar displayed weather information, only the NEXRAD displayed weather information and when both sources of information displayed weather. A third within-groups independent variable, pilot flying, consisted of two levels, situations when the captain was the pilot flying and those when the first officer was the pilot flying.

The dependent variables in this study were deviation decision, deviation accuracy (we assessed whether or not the deviation decisions made by the flight crews were correct or incorrect, compared to criteria specified by expert pilots), pilot confidence in their deviation decision, perceived situation awareness, perceived workload, and trust in both the onboard and NEXRAD weather sources. Additional dichotomous dependent variables were measured concerning pilot teamwork. This first was leadership style. Leadership style could assume one of two possibilities based on Normative Decision Theory: participative or autocratic leadership style. The second dependent variable was

communication level. It assumed one of two possibilities: high and low communication between flight crewmembers.

Participants

The researchers collected data from 15 aviator teams (30 individual aviators) from six airlines, though the majority came from United Airlines (see Figure 1). Data from three teams were not analyzed due to design errors. Specifically, the flight simulator and the weather display program were not properly synchronized to ensure that the weather information was displayed at the proper location along the flight path. The remaining participants were 12 male pilot teams (twelve of Captain rank and 12 of First Officer rank). Female pilots were excluded from the study to allow clearer generalization to the largely male-dominated cockpit environment and to control for possible sex-related interpersonal team effects. Participants were recruited from the Eastern US region through an existing agreement with NASA Langley Research Center, Lockheed Martin, and SWALES Corporation, and were compensated in exchange for their participation (an hourly stipend plus reimbursement of travel expenses).

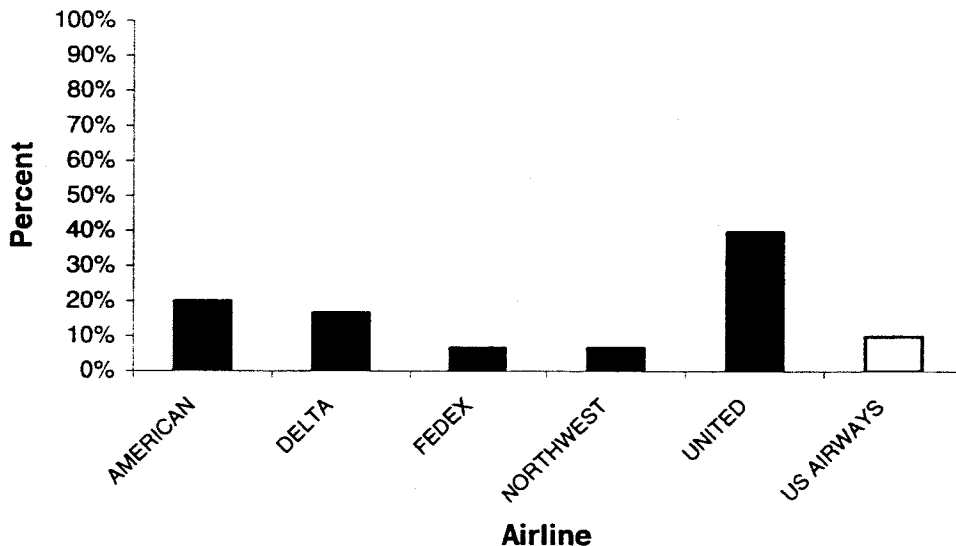


Figure 1. Corporate Representations of Aviator Participants.

Data from two background questionnaires administered prior to the experiment revealed that Captains' ages ranged from 46 to 60 years ($M = 55.13$, $SD = 4.21$), whereas First Officers' ages ranged from 34 to 56 years ($M = 46.33$, $SD = 5.79$). The number of reported hours of glass cockpit experience ranged from 1100 to 1,2000 hours, and the number of pilot flight hours ranged from 5000 to 19000 hours of experience (see Table 1). Only one pilot had his last FAA check ride before 2003 and the majority of pilots (71%) had their last check ride in 2004 (see Table 2). Of the 24 pilots, 16 reported having interacted with an integrated weather display and only 4 pilots reported having flown with their teammate prior to the study. On average, participants reported using a computer 13.63 hours per week and reported playing video games only 33 minutes per week. Seven pilots reported participating in teamed sporting activities at least once a

week. Only one pilot reported participating in any musical group activities, but seven pilots reported participating in team activities unrelated to sports and music.

Table 1. Participants' Experience Characteristics (N = 24).

Characteristic	<i>M</i>	<i>SD</i>
Age at time of study (years)	50.73	5
Glass cockpit experience (hours)	4981.25	2475.39
Flight experience (hours)	11398.34	2582.85
Computer use (hours per week)	13.63	11.86
Video game use (hours per week)	.33	.87

Table 2. Participants' Demographic Characteristics (N = 24).

Characteristic	<i>n</i>	%
Last FAA check ride		
July 2002 - December 2002	1	4
January 2003 - June 2003	1	4
July 2003 - December 2003	5	21
January 2004 - June 2004	15	63
July 2004 - December 2004	2	8
How well do you know the other pilot?		
Never met him before today	17	71
Barely know him (never flown together)	3	13
Know him fairly well (flown together 1-5 times)	4	17
How many times per week do you engage in team sporting activities?		
0	17	71
1	5	21
2	1	4
4	1	4

Materials

The laboratory space used for this study housed three computers. One computer hosted Microsoft Flight Simulator 2004 and was physically connected to the Rudder Control Module, Sub Panel Assembly, external power quadrants and avionics stacks of the EPIC AV-B/IFR General Aviation Flight Console. The flight console came equipped with a flight yoke and basic flight instruments. A second computer to the right of the

flight simulator hosted a Visual Basic 6.0 program, which periodically displayed two sources of weather information throughout the course of the flight (See Appendices 1 and 2). One source of weather information was a static image of the onboard weather radar, constructed in Powerpoint and displayed by the Visual Basic program. The other source was a static image of Next Generation Radar (NEXRAD) weather imagery. The NEXRAD imagery was obtained from the National Environmental Satellite, Data and Information Service, converted to an image file and presented by the Visual Basic program. The computer also presented a series of questions concerning how and whether the flight crew wanted to deviate from the weather. Three of the questions asked the team to rate their confidence on a 0-100 point scale. For example, one question asked participants to rate their level of confidence that a deviation should be made from the upcoming weather event. A fourth question specifically asked the pilots if they would choose to deviate at this time and in what lateral direction i.e. left or right (see Appendix 2). This computer also hosted a background questionnaire, an electronic version of the Situation Awareness Rating Tool (SART - Taylor, 1990)(Appendix 3), an electronic version of the NASA Task Load Index (TLX) workload rating form (Appendix 4), and an electronic 10-item trust questionnaire created by the researchers (Appendix 5). The trust questionnaire was designed to assess pilot trust in the two sources of weather information.

A third computer was located on a 90-degree angle to the left of the flight simulator. This computer also hosted the trust, workload and situation awareness questionnaires, for completion by the pilot flying. All the computers had Intel Pentium IV processors and flat screen 17-inch monitors. The pilots completed all computerized questionnaires using a standard QWERTY keyboard and mouse.

Prior to each flight leg, the pilots also received preflight briefing information (See Appendix 6). This information included a graphical depiction of the flight path and a packet of weather information. The weather packet included general information such as wind speed, direction and convective activity in the United States. The usefulness of this information was limited by its age. The pilots were informed that this information was 8 hours old.

Participants also completed paper background and opinion questionnaires (see Appendices 7 and 8). The background questionnaire was designed to obtain pertinent background information, such as age and amount of glass cockpit experience. The opinion questionnaire contained 5-point Likert scale items designed to reveal pilot strategies used for performing the task and pilot opinion of information quality. For example, one item asked pilots to rate the realism of the weather presentation system. Finally, pilots were videotaped periodically throughout the study using a Sony 8mm Video Camera/Recorder.

A research proposal was submitted to the Old Dominion University Institutional Review Board (IRB) to insure that the research protocol conformed to the American Psychological Association's ethical guidelines. In addition, all participants were required to complete an informed consent form prior to participation (see Appendix 9). Also, the participants were properly debriefed after each experimental session and received experimenter contact information if they had any future questions regarding the nature of the study.

Procedure

Using an existing agreement between NASA Langley Research Center, Lockheed Martin, and SWALES Corporation, aviator participants were recruited specifically for participation in this research, and were compensated for their participation. When the pilots arrived, they received an informed consent form to read and sign. Next the experimenter administered the paper participant background questionnaire and randomly assigned the members of the flight crew to one of the two seating arrangements (captain as pilot flying, or first officer as pilot flying), to maintain a true experimental design (Tabachnick & Fidell, 2001). The aviator assigned to the pilot flying seat sat in front of the computer hosting Microsoft Flight Simulator and the pilot assigned to pilot not flying seat was instructed to sit at the computer which displayed the weather information.

At this point the pilots worked together to complete the computerized background questionnaire on the pilot not flying computer screen. Next, the experimenter provided the team with a brief overview of the study (Appendix 10) and administered written instructions to aid the pilots on the proper completion of the NASA TLX, SART and trust questionnaires. The pilots read through the instructions and were advised to refer to them throughout the study.

To familiarize the pilots and reduce practice effects the pilots were instructed to first fly a practice flight leg from Sacramento, CA to Los Angeles, CA. Before the flight the experimenter administered the preflight briefing information. After reading through this information the pilots began the practice flight. To properly begin the simulated flight, the pilots were instructed to begin the flight simulator and visual basic weather display program simultaneously. Beginning both programs at the same time was important to properly synchronize the location of the weather information on the display computer with the team's location on the simulated flight path. The pilots were not required to take off or land the simulator; the flight began in the air at an altitude of 19000 feet. Participants were instructed to maintain this altitude, and an airspeed of 325 nautical miles per hour.

During most of the flight the weather display computer did not display any information on the monitor. The program would display weather information only at set distances from weather events. During the practice session pilots encountered only one potential weather event. At 160 nautical miles into the flight, the two weather displays and a series of deviation questions flashed on the weather display monitor. At this point, the Captain was instructed to disengage the autopilot and fly the plane manually. In addition the experimenter began video recording the subsequent team interaction. The Captain and First Officer worked as a team to complete the series of deviation questions based on the two sources of weather information. Although the pilots were permitted to work together, they were reminded that the Captain was to give final approval of any deviation decision that was reached. Pilots were allotted 3.5 minutes to answer the four deviation questions. When the pilots completed their deviation decision and questionnaires the experimenter stopped recording.

After completion of the deviation questions the pilot flying was instructed to pause the flight simulator and both team members completed the NASA TLX, SART and trust questionnaires independently on separate computers. The pilot flying completed his questionnaire on the computer located 90 degrees to his left and the pilot not flying

completed these measures on the weather display computer. The pilots completed two trust questionnaires, one for each source of weather information.

Once the pilots completed their computerized questionnaires the pilot flying took his position at the flight simulator and reengaged the simulator. Next, the team reengaged the autopilot and weather display program simultaneously and continued along the flight path. The pilots were not permitted to actually deviate from the flight path. As the team approached the weather event they received three more presentations of the weather at 80, 40 and 20 nautical miles from the event. The same procedure was followed for every presentation.

After the practice flight leg the pilots took a ten-minute break during which time the experimenter answered any questions. After the break, the participants began the experimental flight legs. The experimental procedure was identical to the practice procedure, except that the flight crews encountered three weather events per flight leg. This allowed presentation of each combination of levels of the three independent variables.

The first experimental flight leg was a flight from New York's John F. Kennedy Airport to Miami's International Airport. The flight took approximately 3.5 hours, after which the pilots were provided with a 1-hour break for lunch. After lunch, they returned to the experimental laboratory to complete the second experimental flight leg (a return leg from Miami, FL back to New York). The Captain and First Officer switched seats for the return leg, so that each aviator was given the chance to fly the simulator.

Once the experimental flights were complete the pilots were instructed to complete the opinion questionnaire. The pilots were then orally debriefed and thanked for their participation. The study took approximately 7.5 hours to complete.

RESULTS

Background Information

In addition to the descriptive information presented earlier, we calculated several comparisons of demographic information between those participants who were captains and those who were first officers.

Age. An independent-samples t-test showed a statistically significant difference in age between Captains and First Officers, $t(28) = 4.76, p < .001$. On average, Captains ($M = 55.13, SD = 4.21$) were significantly older than First Officers ($M = 46.33, SD = 5.79$).

Flight hours. An independent-samples t-test also showed a statistically significant difference in flight hours between Captains and First Officers, $t(28) = 4.78, p < .001$. On average, Captains had a significantly greater number of flight hours ($M = 13666.67, SD = 2888.81$) than First Officers ($M = 9130, SD = 2276.89$).

Glass Cockpit and Integrated Weather Display Experience. A bivariate correlation showed a significant negative correlation for hours of glass cockpit experience and whether or not pilots have interacted with an integrated weather display, $r(24) = -.55, p < .01$. The more glass cockpit experience a pilot had the less likely they were to have interacted with an integrated weather display. An independent-samples t-

test showed no significant difference between Pilots and First Officers for amount of glass cockpit experience ($p > .05$).

Weather Confidence Ratings

Confidence that the upcoming event actually existed. We examined the effects of pilot flying, weather display agreement, and distance to weather on teams' confidence that the upcoming weather event actually existed through a 2X3X4 repeated-measures ANOVA. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as independent variables. Teams' confidence that the upcoming weather event actually existed was used as the dependent variable. Results showed a statistically significant two-way interaction effect of Pilot Flying and Distance, $F(3, 33) = 3.72, p < .05$, partial $\eta^2 = .25$. Results also showed a statistically significant main effect of Distance, $F(3, 33) = 3.56, p < .05$, partial $\eta^2 = .25$. Follow-up pairwise comparisons showed that teams' confidence when the captain was flying ($M = 80.92, SD = 25.55$) was significantly greater than when the first officer was flying ($M = 71.36, SD = 29.35$) at 160 nm. Similarly, teams' confidence when the captain was flying ($M = 91.11, SD = 11.62$) was significantly greater than when the first officer was flying ($M = 80.00, SD = 30.61$) at 20 nm. However, teams' confidence when the captain was flying was similar to when the first officer was flying at 80 nm and 40 nm. These results are graphically depicted in Figure 2.

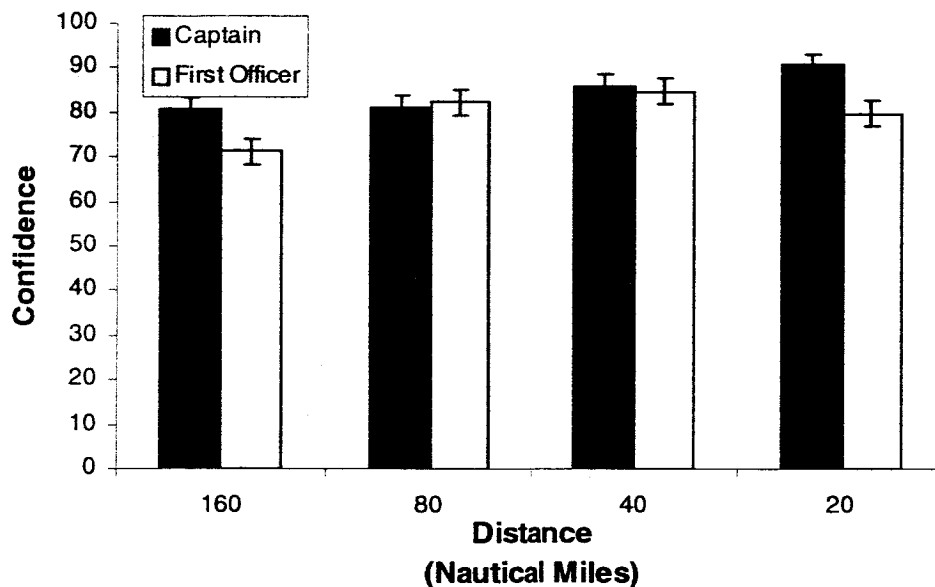


Figure 2. Pilot Confidence Ratings as a Function of Distance to the Weather Event.

Results also showed a statistically significant main effect of Agreement, $F(1.14, 12.58) = 9.91, p < .01$, partial $\eta^2 = .47$. Follow-up pairwise comparisons showed that teams' confidence when both systems agreed ($M = 91.83, SD = 9.81$) was significantly greater than when only the NEXRAD system indicated that there was an upcoming weather event ($M = 68.11, SD = 34.04$). These results are depicted in Figure 3.

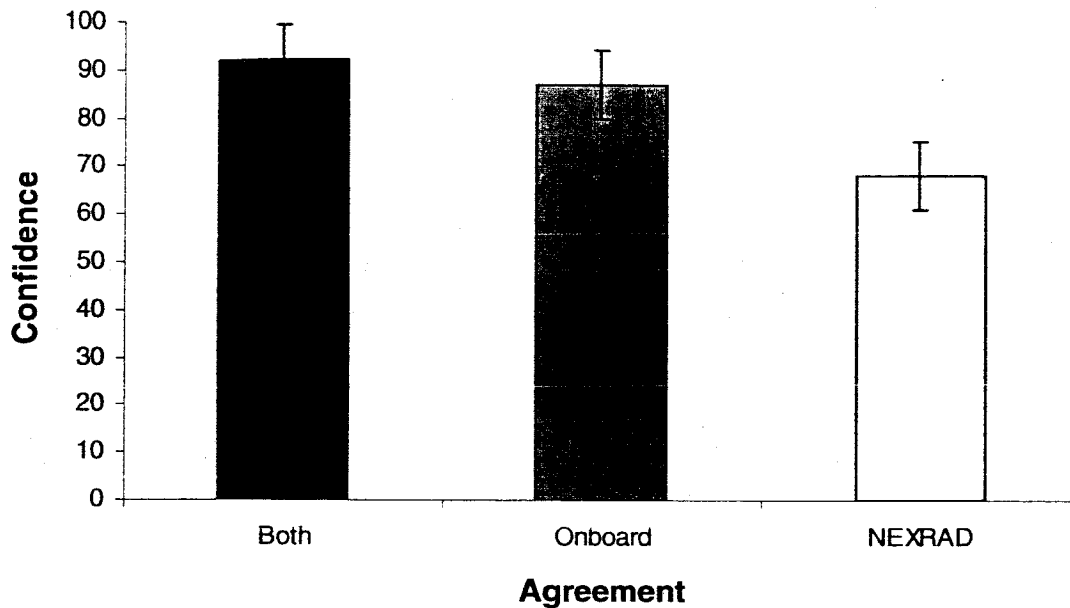


Figure 3. Decision Confidence Level as a Function of Display Agreement.

Confidence that flight crew should deviate. We examined the effects of pilot flying, systems' agreement, and distance to weather on teams' confidence that they should deviate through a 2X3X4 repeated-measures ANOVA. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as independent variables. Teams' confidence that they should deviate was used as the dependent variable. Results showed a statistically significant two-way interaction effect of Agreement and Distance, $F(2.76, 30.38) = 6.86, p < .01, \text{partial } \eta^2 = .38$. Results also showed statistically significant main effects of Agreement, $F(1.36, 14.96) = 52.13, p < .001, \text{partial } \eta^2 = .83$, and Distance, $F(1.22, 13.41) = 22.13, p < .001, \text{partial } \eta^2 = .67$. Simple effect follow-ups showed that confidence significantly improved as a function of distance when both systems agreed, $F(1.07, 24.68) = 26.68, p < .001, \text{partial } \eta^2 = .54$, and when only the Onboard system indicated that there was an upcoming weather event, $F(1.16, 26.69) = 35.15, p < .001, \text{partial } \eta^2 = .60$, but not when only the NEXRAD system indicated that there was an upcoming weather event, $F(2.12, 48.84) = .37, \text{n.s.}, \text{partial } \eta^2 = .02$. These results are graphically depicted in Figure 4.

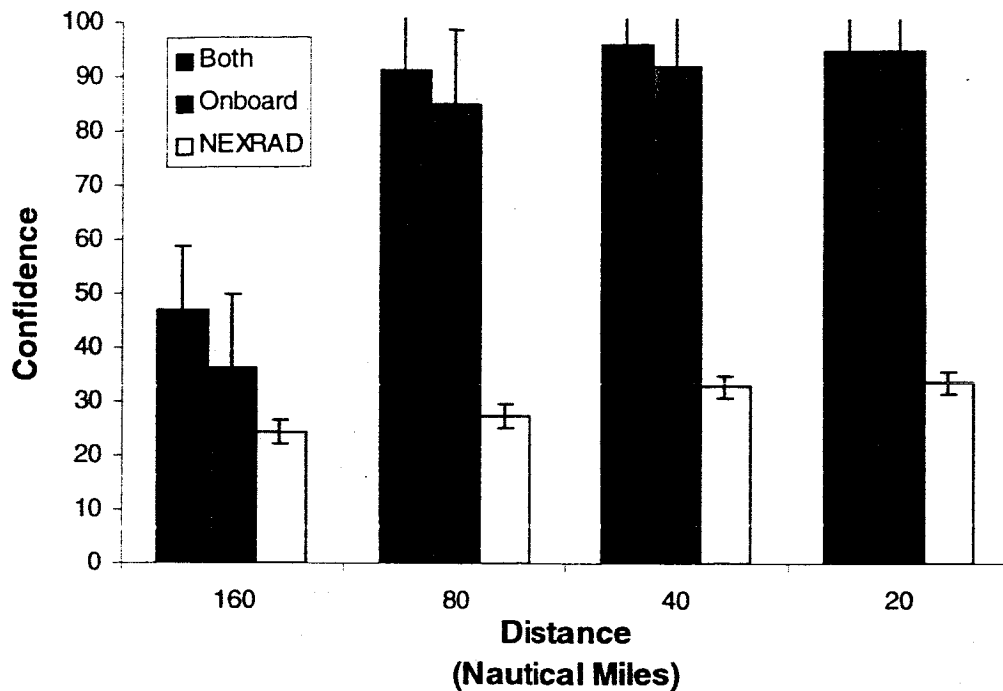


Figure 4. Confidence Levels as a Function of Distance to the Weather Event and Weather Event Display.

Trust

Because the trust questionnaire we used had not been used for aviation research previously, we calculated its internal consistency reliability. The resulting alpha value was .98, suggesting that the questionnaire had excellent internal consistency.

We examined the effects of pilot, pilot flying, system, systems' agreement, and distance on pilots' trust through a 2X2X2X3X4 mixed ANOVA. Pilot (Captain, First Officer) was used as the between-groups independent variable. Pilot Flying (Captain, First Officer), System (NEXRAD, Onboard), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as within-groups independent variables. Pilots' trust was used as the dependent variable. Results showed a statistically significant three-way interaction effect of System, Agreement, and Distance, $F(3.12, 69.45) = 9.82, p < .001, \text{partial } \eta^2 = .31$. Results also showed statistically significant two-way interaction effects of System and Agreement, $F(1.20, 26.36) = 35.54, p < .001, \text{partial } \eta^2 = .62$, System and Distance, $F(2.12, 46.53) = 3.46, p < .05, \text{partial } \eta^2 = .14$, and Agreement and Distance, $F(3.80, 83.70) = 5.44, p < .01, \text{partial } \eta^2 = .20$. Lastly, results showed statistically significant main effects of System, $F(1, 22) = 37.31, p < .001, \text{partial } \eta^2 = .63$, Agreement, $F(2, 40.25) = 16.90, p < .001, \text{partial } \eta^2 = .43$, and Distance, $F(3, 66) = 4.88, p < .01, \text{partial } \eta^2 = .18$.

Simple effect follow-ups showed that pilots' trust of the NEXRAD system significantly decreased as distance to the weather decreased when the NEXRAD system did not provide them with an indication of an upcoming weather event, $F(1.59, 74.66) = 20.49, p < .001, \text{partial } \eta^2 = .30$. Trust also significantly decreased as a function of

distance when the onboard weather display did not provide them with an indication of an upcoming weather event, $F(2.02, 94.98) = 5.37, p < .01$, partial $\eta^2 = .10$. On the other hand, pilots' trust on the Onboard system significantly increased as a function of distance when both systems agreed, $F(2.16, 101.60) = 6.12, p < .01$, partial $\eta^2 = .12$, and when it was the only system that indicated the presence of an upcoming weather event, $F(1.99, 93.39) = 3.49, p < .05$, partial $\eta^2 = .07$. These results are graphically depicted in Figures 5 and 6.

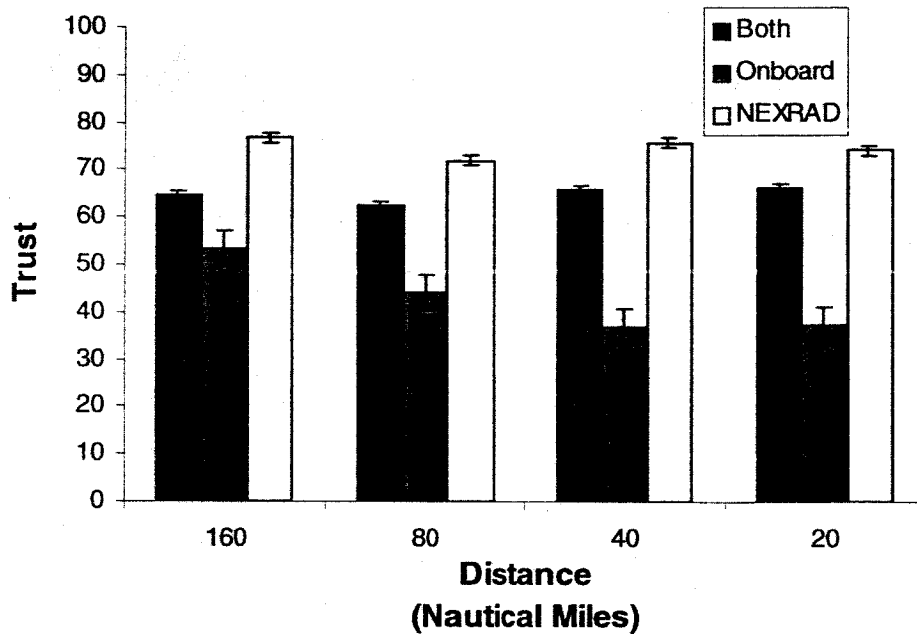


Figure 5. Pilot Flying Trust in the Weather Display as a Function of Distance to the Weather Event and Display Agreement.

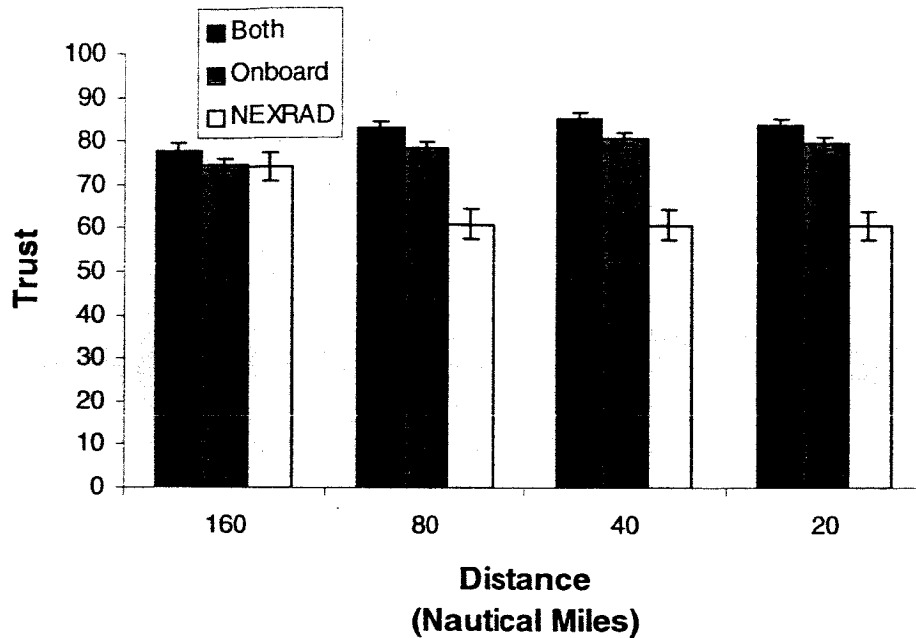


Figure 6. Pilot Not Flying Trust in the Weather Event as a Function of Distance to the Weather Event and Display Agreement.

Perceived Workload

We examined the effects of pilot, pilot flying, systems' agreement, and distance on pilots' perceived workload through a 2X2X2X3X4 mixed ANOVA. Pilot (Captain, First Officer) was used as the between-groups independent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as within-groups independent variables. Pilots' perceived workload was used as the dependent variable. Results showed a statistically significant main effect of Distance, $F(3,66) = 8.33, p < .001$, partial $\eta^2 = .28$. Follow-up pairwise comparisons showed that pilots' perceived workload significantly increased as distance decreased from 160 nm ($M = 26.45, SD = 17.98$) to 20 nm ($M = 29.76, SD = 18.18$). These results are graphically depicted in Figure 7.

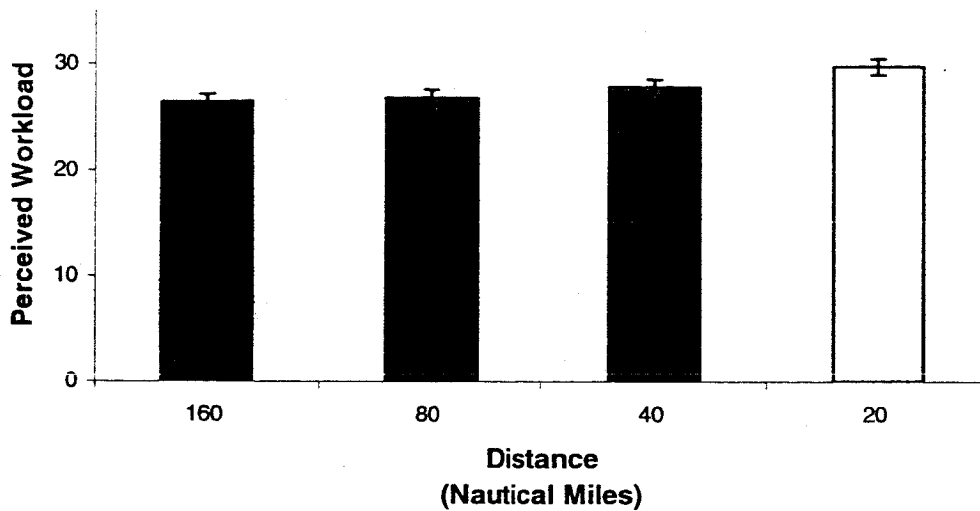


Figure 7. Perceived Workload as a Function of Distance to the Weather Event.

Perceived Situation Awareness

We examined the effects of pilot, pilot flying, systems' agreement, and distance on pilots' perceived situation awareness through a 2X2X2X3X4 mixed ANOVA. Pilot (Captain, First Officer) was used as the between-groups independent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as within-groups independent variables. Pilots' perceived workload was used as the dependent variable. Results showed a statistically significant two-way interaction effect of Agreement and Distance, $F(3.08, 67.68) = 2.81, p < .05$, partial $\eta^2 = .11$. Results also showed significant main effects of Agreement, $F(1.32, 29.09) = 5.57, p < .01$, partial $\eta^2 = .20$, and Distance, $F(1.95, 42.83) = 7.90, p < .01$, partial $\eta^2 = .26$. Simple effect follow-ups showed that pilots' perceived situation awareness significantly decreased as a function of distance when the NEXRAD system did not provide them with an indication of an upcoming weather event, $F(1.93, 90.59) = 7.56, p < .01$, partial $\eta^2 = .14$. These results are graphically depicted in Figure 8.

Results also showed a statistically significant main effect of Pilot, $F(1, 22) = 5.34, p < .05$, partial $\eta^2 = .20$. First officers reported a significantly higher level of situation awareness ($M = 32.10, SD = 7.73$), than Captains ($M = 26.15, SD = 7.20$). These results are graphically depicted in Figure 9.

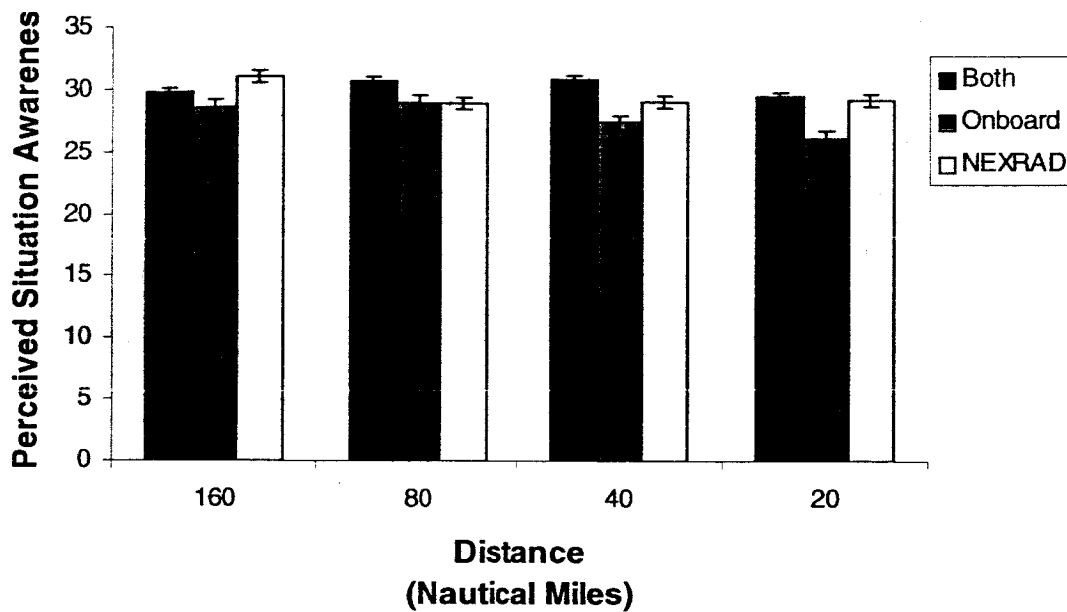


Figure 8. Perceived Situation Awareness as a Function of Distance to the Weather Event and Display Agreement.

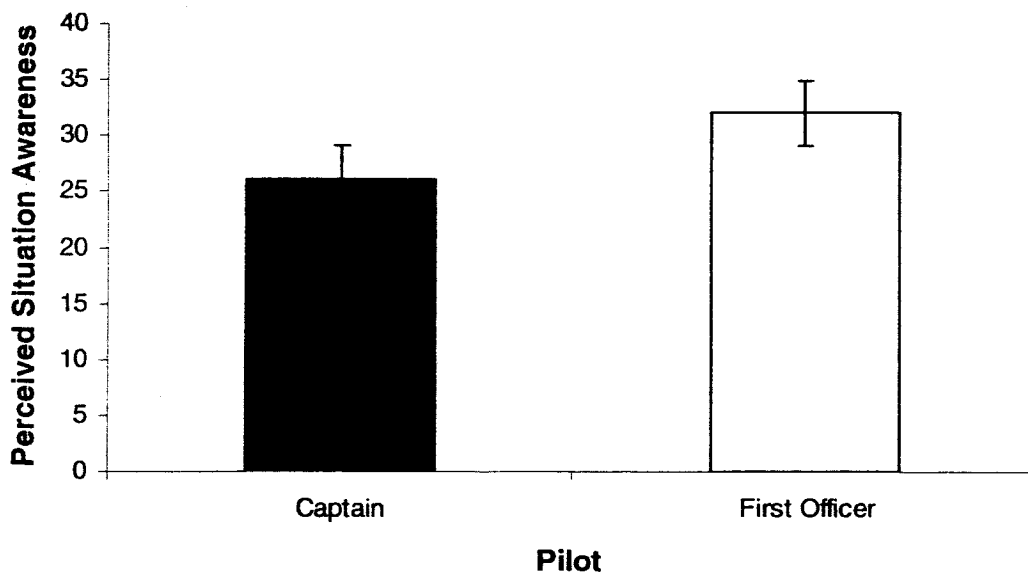


Figure 9. Perceived Situation Awareness as a Function of Pilot Classification.

Deviation Decision

A Chi-Square test showed that teams were significantly more likely to want to deviate from the flight path than stay on course, $\chi^2(1) = 28.13, p < .001$. Of the 288 deviation decisions made, teams wanted to deviate from the flight path 189 times. We

examined all the variables that could predict teams' decision weather to deviate from the flight path or stay in course through a standard logistic regression.

Teams' deviation decision (No, Yes) was used as the dependent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), Distance (160 nm, 80 nm, 40 nm, 20 nm), teams' confidence that the upcoming weather event actually existed, teams' confidence that they should deviate, teams' confidence about their decision, pilots' trust on the onboard and NEXRAD systems, pilots' perceived workload, and pilots' perceived situation awareness were used as predictors.

Results from the standard logistic regression indicated that the combination of the predictors listed earlier significantly predicted the outcome, $\chi^2(17) = 313.16, p < .001, R^2 = .66$. However, results from each individual Wald statistic indicated that only agreement, distance, and teams' confidence that they should deviate were significant predictors of their deviation decision. Therefore, we conducted a follow-up standard logistic regression that included only these three predictors. Results from this analysis indicated that the combination of just these three predictors significantly predicted the outcome, $\chi^2(6) = 292.81, p < .001, R^2 = .64$. A total of 95.1% of all teams' decisions were correctly predicted with this model. Type I error was 2.6%, indicating that 97.4% of teams' decisions to want to deviate from the flight path were correctly classified. Type II error was 9.1%, indicating that 90.9% of teams' decisions to want to stay in course were correctly classified. Teams were .08 times less likely to want to deviate from the flight path when the onboard system did not provide them with an indication of an upcoming weather event. Also, teams were 21.58 times more likely to want deviate from the flight path when they were 40 nm away from the upcoming weather event. Ultimately, teams were 1.07 times more likely to want to deviate from the flight path with every unit increase in their confidence that they should deviate. These results are summarized in Table 3.

Table 3. Standard Logistic Regression to Predict Deviation Decision.

Variable	B	SE	Wald statistic	Odds Ratio ¹
Agreement			7.23*	
Onboard	-1.16	1.00	1.35	.32 (.05 to 2.22)
NEXRAD	-2.51	.98	6.55*	.08 (.01 to .55)
Distance			12.16 **	
80 nm	2.84	1.01	7.86**	17.06 (2.35 to 124.06)
40 nm	3.07	1.03	8.83**	21.58 (2.85 to 163.68)
20 nm	1.88	.81	5.31*	6.52 (1.32 to 32.13)
Confidence	.07	.01	48.87 ***	1.07 (1.06 to 1.10)

¹ Confidence intervals in parentheses; * $p < .05$, ** $p < .01$, *** $p < .001$

Weather Confidence Ratings

Confidence in their Decision. We examined the effects of pilot flying, systems' agreement, and distance to weather on teams' confidence in their decision using a 2X3X4 repeated-measures ANOVA. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), and Distance (160 nm, 80 nm, 40 nm, 20 nm) were used as independent variables. Teams' confidence in their decision was used as the dependent variable. Results showed a significant main effect of Agreement, $F(2, 22) = 3.35$, $p = .05$, partial $\eta^2 = .23$. Teams' confidence in their decision was highest when both systems agreed ($M = 92.66$), followed by when only the onboard system indicated the presence of an upcoming weather event ($M = 90.39$) and when only the NEXRAD system indicated the presence of an upcoming weather event ($M = 86.64$). These results are graphically depicted in Figure 10.

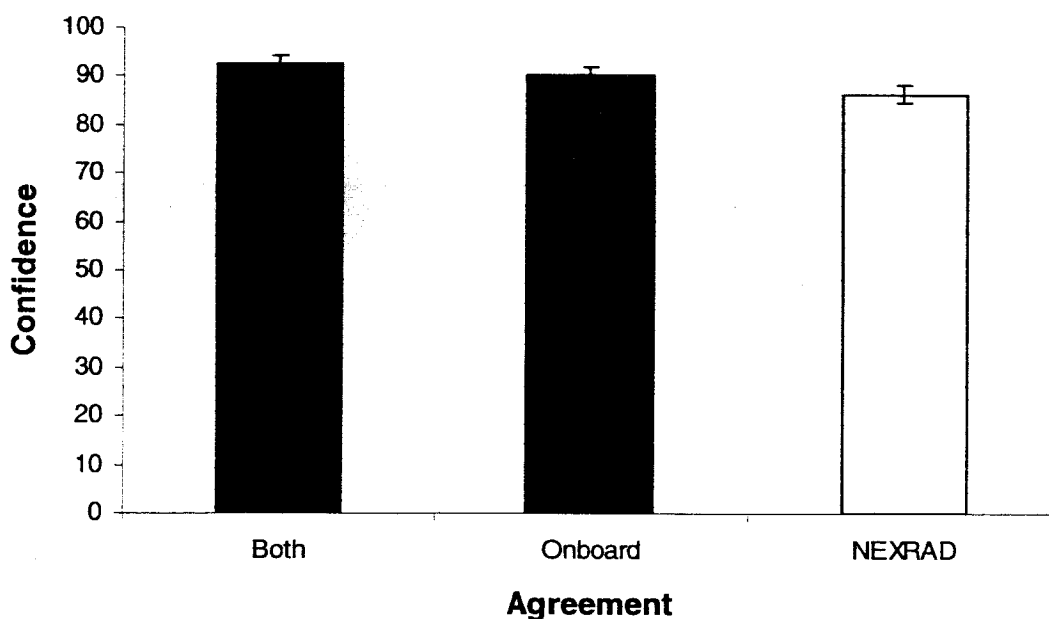


Figure 10. Confidence in the Deviation Decision as a Function of Weather System Agreement.

Videotaped Recordings

The research team designed a rating system to use when analyzing the videotaped crew performances. That system allowed researchers to assess crew communication levels and leadership style (see Appendix 11). Two raters used the rating system to review the videotaped experimental sessions and assess each crew's leadership style and level of communication at 160, 80, 40, and 20 miles from each of the six weather events. Level of communication referred to the amount of discussion concerning the

interpretation of and responses to the weather information. It was rated on two levels (1 = low/infrequent communication, 2 = high/frequent communication). Leadership style was assessed according to the characteristics of authoritative and participative leadership styles defined by the Normative Decision Theory (Chemers, 2000).

Two raters independently reviewed the videotaped sessions. The videotapes were viewed in sequential order from Team 3 to Team 15, omitting Teams 1, 2, and 4 because of problems associated with the audio recording quality and flight segment synchronization. At each decision point, the raters provided a score that reflected the participants' level of communication (1 = low/infrequent communication; 2=high/frequent communication). In addition to the communication scores, the raters identified the leadership style employed by the crews at each decision point. The leadership style was participative or authoritative based on Normative Decision Theory. Raters were allowed to rewind and review pilot-copilot interactions as often as they deemed necessary.

The rating system was developed solely for the purposes of this research; therefore, its efficacy had not been determined prior to use. For that reason, inter-rater reliability was calculated after the ratings were compiled, prior to performing any subsequent data analyses. Initial inter-rater reliability coefficients revealed only moderate agreement between the raters for communication level ($\Phi = .46$) and leadership style (.34). To allow analyses of the data, the raters subsequently met and discussed the differences in the ratings and came to agreement regarding them. Appendix 12 shows the resulting consensus ratings.

Leadership Style

A Chi-Square test showed that teams were significantly more likely to use a participative leadership style than an autocratic leadership style, $\chi^2(1) = 84.5, p < .001$. Of the 288 deviation decisions made, teams used a participative leadership style 222 times. We examined all the variables that could predict teams' leadership style through a standard logistic regression. Teams' leadership style (Participative, Autocratic) was used as the dependent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), Distance (160 nm, 80 nm, 40 nm, 20 nm), pilots' age, and pilots' flight hours were used as independent variables. Results from the standard logistic regression indicated that the combination of these predictors significantly predicted teams' leadership style, $\chi^2(10) = 76.84, p < .001, R^2 = .23$. However, results from each individual Wald statistic indicated that only the captains' age was a significant predictor of leadership style. Therefore, we conducted a follow-up standard logistic regression including just this predictor. Results from this analysis indicated that captains' age significantly predicted teams' leadership style, $\chi^2(1) = 60.11, p < .001, R^2 = .19$. A total of 81.90% of all teams' leadership styles were correctly predicted with this model. Type I error was 7.70%, indicating that 92.30% of teams' leadership styles were correctly classified as participative. Type II error was 53.00%, indicating that 47% of teams' leadership styles were correctly classified as autocratic. Teams were .75 times less likely to use an autocratic leadership style with every unit increase in captains' age. These results are summarized in Table 4.

Table 4. Logistic Regression Examining Leadership Style.

Variable	B	SE	Wald statistic	Odds Ratio ¹
Age (CAPT)	-.28	.04	47.95***	.75 (.70 to .82)

¹ Confidence intervals in parentheses, *** $p < .001$

Communication

A Chi-Square test showed no significant differences in the ratings of teams' communication, $\chi^2(1) = .01, p = .91$. Of the 288 possible decision points where communication was rated, teams were rated as exhibiting a low level of communication 145 times and a high level of communication 143 times. We examined all the variables that could predict teams' communication through a standard logistic regression. Teams' Communication (low, high) was used as the dependent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), Distance (160 nm, 80 nm, 40 nm, 20 nm), Leadership (Participative, Autocratic), pilots' age, pilots' flight hours, teams' confidence that the upcoming weather event actually existed, pilots' trust on the onboard and NEXRAD systems, pilots' perceived workload, and pilots' perceived situation awareness were used as predictors.

Results from the standard logistic regression indicated that the combination of these predictors significantly predicted the outcome, $\chi^2(20) = 87.42, p < .001, R^2 = .26$. However, results from each individual Wald statistic indicated that only Leadership, captains' age, captains' flight hours, and captains' perceived situation awareness were significant predictors of communication level. Therefore, we conducted a follow-up standard logistic regression including only these four predictors. Results from this analysis indicated that the combination of these predictors significantly predicted the outcome, $\chi^2(4) = 73.24, p < .001, R^2 = .23$. A total of 71.90% of all teams' decisions were correctly predicted with this model. Type I error was 19.60%, indicating that 80.40% of teams' high communication levels were correctly classified. Type II error was 36.60%, indicating that 63.40% of teams' low communication levels were correctly classified. Teams were .21 times less likely to exhibit a high level of communication when they were classified as having an autocratic leadership style. Also, teams were 1.10 times more likely to exhibit a high level of communication with every unit increase in the captains' age. In addition, teams were 1.00 times more likely to exhibit a high level of communication with every unit increase in the captains' flight hours. Finally, teams were 1.05 times more likely to exhibit a high level of communication with every unit increase in the captains' perceived situational awareness. These results are summarized in Table 5.

Table 5. Logistic Regression Examining Communication.

Variable	B	SE	Wald statistic	Odds Ratio ¹
----------	---	----	----------------	-------------------------

Leadership	-1.54	.40	14.69***	.21 (.10 to .47)
Age (CAPT)	.10	.04	4.95*	1.10 (1.01 to 1.20)
Flight Hours (CAPT)	.00	.00	12.42***	1.00 (1.00 to 1.00)
Perceived SA (CAPT)	.05	.02	6.20*	1.05 (1.01 to 1.10)

¹ Confidence intervals in parentheses, * $p < .05$; *** $p < .001$

Accuracy

To determine whether crews had made the optimal deviation decision at each of the weather decision points, the experimenters arranged for two current air transport captains to collaboratively review the flight path and tasks. After doing so, these subject matter experts provided a scoring key against which the experimental participants' decisions could be compared. The subject matter experts recommended deviation decisions for each decision point, according to three ranked criteria: safety of the flight (most important), comfort of the passengers, and economy of the flight (least important).

Subsequent to comparing the experimental results to the scoring key, a Chi-Square test showed that teams were significantly more likely to make an accurate deviation decision than an inaccurate deviation decision, $\chi^2(1) = 10.13, p < .01$. Of the 288 possible deviation decisions, teams made 171 accurate deviation decisions. Experimenters examined all the variables that could predict teams' deviation decision accuracy through a standard logistic regression. Teams' deviation decision accuracy was used as the dependent variable. Pilot Flying (Captain, First Officer), Agreement (Both, Onboard, NEXRAD), Distance (160 nm, 80 nm, 40 nm, 20 nm), Communication (Low, High), Leadership (Participative, Autocratic), pilots' age, pilots' flight hours, teams' confidence in their deviation decision, pilots' trust on the onboard and NEXRAD systems, pilots' perceived workload, and pilots' perceived situation awareness were used as predictors. Results from the standard logistic regression indicated that the combination of these predictors significantly predicted deviation decision accuracy, $\chi^2(21) = 73.19, p < .001, R^2 = .22$. However, results from each individual Wald statistic indicated that only agreement, distance, communication, and pilots' trust in the onboard system were significant predictors of teams' deviation decision accuracy. Therefore, we conducted a follow-up standard logistic regression including just these five predictors.

Results from this analysis indicated that the combination of just these five predictors significantly predicted deviation decision accuracy, $\chi^2(8) = 61.47, p < .001, R^2 = .19$. A total of 70.50% of all teams' deviation decisions were correctly predicted with this model. Type I error was 42.70%, indicating that 57.30% of teams' inaccurate deviation decisions were correctly classified as inaccurate. Type II error was 20.50%, indicating that 79.50% of teams' accurate deviation decisions were correctly classified as accurate. Teams were .49 times less likely to make an accurate deviation decision when only the onboard system indicated an upcoming weather event. Teams were 4.01 times more likely to make an accurate deviation decision when only the NEXRAD system indicated an upcoming weather event. Teams were 3.76 times more likely to make an accurate deviation decision when they were 20 nm away from the upcoming weather event. Teams were .51 times less likely to make an accurate deviation decision when they

were rated as having a high communication level. Teams were 1.02 times more likely to make an accurate deviation decision with every unit increase of their trust in the onboard system. These results are summarized in Table 6.

Table 6. Logistic Regression Examining Deviation Decision Accuracy.

Variable	B	SE	Wald statistic	Odds Ratio ¹
Agreement			29.13***	
Onboard	-.71	.32	5.06*	.49 (.26 to .91)
NEXRAD	1.41	.41	12.11**	4.10 (1.85 to 9.06)
Distance			13.55 **	
80 nm	.87	.38	5.21*	2.38 (1.13 to 5.00)
40 nm	.29	.37	.61	1.34 (.65 to 2.76)
20 nm	1.32	.40	11.19**	3.76 (1.73 to 8.17)
Communication	-.68	.27	6.16*	.51 (.30 to .87)
Trust in Onboard (CAPT)	.03	.01	8.80**	1.03 (1.01 to 1.05)
Trust in Onboard (FO)	.02	.01	5.39*	1.02 (1.00 to 1.04)

¹ Confidence intervals in parentheses, * $p < .05$; ** $p < .01$; *** $p < .001$

Opinion Questionnaire

Items #4 and #5 of the Opinion Questionnaire asked participants to retrospectively rate their levels of situation awareness and workload, respectively (See Appendix 8). An independent samples t-test was conducted to examine the effects of pilot rank (Captain or First Officer) on these one-item measures of perceived workload and perceived situation awareness. The t-tests showed no significant difference between Captains and First Officers ($p > .05$).

Mean scores for the SART and NASA TLX were computed across all conditions for each pilot. Bivariate correlations were computed to assess the correlations between these scores and the retrospective scores from the opinion questionnaire assessing situation awareness and workload. The NASA TLX was significantly positively correlated with the opinion questionnaire workload measure, $r(24) = .55, p < .01$. However, the SART was not significantly correlated with the opinion questionnaire situation awareness measure ($p > .05$). These results are summarized in Table 7.

Table 7. Correlations Among Situation Awareness and Workload Measures.

Variable	1	2	3
1. NASA TLX	--		
2. SART	-.61*	--	
3. Opinion Questionnaire (WL)	.55*	-.34	--
4. Opinion Questionnaire (SA)	.16	.11	.04

*p<.01

DISCUSSION

Performance-Based Data

The statistical findings presented above reveal interesting patterns with regard to trust of weather displays and the role of interpersonal dynamics on trust, perceived situation awareness and workload. As noted in the introduction to this report, crew reactions to the weather displays presented here may be separated into those that are performance-based and those that are based on subjective impressions of workload, situation awareness, and trust.

Our main performance-based variable, deviation decision accuracy, was shown to be predictable by five factors: onboard/NEXRAD display agreement, distance to the weather event, communication level, and pilots' trust in the onboard weather display. Yet, the magnitude of the prediction was only moderately compelling; 70.5% of deviation decisions were correctly predicted. Such marginal predictability is a testament to the complexity surrounding deviation decisions in operational settings. As remarked by a number of participant crews, many of the deviation decisions would normally be influenced by factors not present in the current study, such as directives by air traffic control, presence and behaviors of other traffic in the area, or inflexibilities associated with the flight timetable. The fact that deviation decisions were predictable to any degree was likely reflective of the tight constraints placed on the flight simulation and the professionalism of the aviator crews who participated in the research.

That being said, the fact that display agreement predicted deviation performance falls in line with established research showing the importance of redundancy in flight displays (Selcon, Taylor, & Shadrake, 1991). It also adds support to the idea that pilots may seek to integrate weather views from additional sources with NEXRAD imagery to help them make deviation decisions (Beringer & Ball, 2004).

Although it is heartening to note that the majority of deviation decisions were correct ones according to the expert-generated key, once again the magnitude of the percentage (171 out of 288, or 59%) is certainly not overwhelming. One reason for this low percentage may be the interdependent nature of the deviation decisions. For each weather event, deviation decisions made at the 160-mile range were highly influential on decisions made at closer ranges. Therefore, if participants made an incorrect decision

160 miles from the weather event, that decision was likely to remain incorrect as the distance dropped. Interestingly, crews were more almost four times more likely to make a correct deviation decision at the 20-mile range than at the farther ranges. Perhaps the rules followed to ensure safety, comfort and economy are more clear cut at short ranges.

Regardless of their accuracy level, it is clear that crews were more confident about their decisions when both weather display systems backed them up. This suggests that participants were indeed cueing on the weather as it was displayed, although their trust levels were never terribly high (see Figures 5 and 6).

In general, crews seemed to work in a fairly conservative manner, opting to deviate more often than ride through the weather. It is not surprising that crews were almost 22 times more likely to deviate as the distance to the weather event closed to 40 nm. However, once crews reached 20 nm, their likelihood of deviation dropped dramatically, perhaps signaling that they had committed themselves to their chosen path. This is an interesting finding, perhaps suggesting that 40 nm may represent a sort of cognitive "point of no return," after which crews are likely to simply ride out the impending weather.

Judgment-Based Data

Trust. It is clear that the majority of our analyses concerned data that were of a judgmental nature. In some ways this is appropriate because of the nature of trust as an attitude. Many researchers have demonstrated the link between mistrust of displayed information and performance (c.f., Bliss, 1993; Breznitz, 1984; Getty, Swets, Pickett, & Gonthier, 1995). Empirical explorations of the construct of trust, however, are less common. An exception to this is the work of Gupta, Bisantz, and Singh (2002), who empirically constructed a trust questionnaire that was used as the basis for the trust questionnaire used in this research. Their questionnaire related the concept of trust to a number of other adjectives. However, the participants they used to develop the questionnaire were not aviators, so the applicability of that questionnaire for the current task was in question. After reviewing Gupta et al.'s original questionnaire, we determined that it would likely require modification to be used for our particular task and participant population. For that reason, we elected to modify the original questionnaire by incorporating slightly different adjectives and descriptors. Doing so, we believe, led to gains in relevance and substantial internal consistency.

In general, the trust levels we observed indicate that flight crews were more likely to assume a conservative reaction philosophy. They were more liable to trust the weather display that showed weather than the one that did not. In addition, as distance to a weather event decreased, participants seemed to progressively lose faith in both the NEXRAD and onboard systems if they did not show weather at lower ranges. This may again suggest pilot skepticism toward displays that did not show weather. In addition to the participants' conservative philosophy, another contributing factor to this might be the low base rate for weather problems in general. In the real world (and in this experiment), significant weather events are relatively infrequent occurrences. Therefore, in this experiment participants may have paid particular attention to potential weather events, even though they were told that the displays were not 100% reliable.

A particularly intriguing finding is that the pattern of trust seemed to vary between the pilot flying and the pilot not flying. When pilots flew the simulator they

apparently placed more faith in the information generated by the NEXRAD system. However, pilots who were not flying trusted the onboard system more. This disparity is difficult to understand, because both pilots were given equal access to both sources of weather information. One possible explanation might be that because the NEXRAD system more clearly showed the full extent of the weather cell, it was more compelling for pilots who were actively in control of the simulator. Conversely, pilots who were not actively in control of the simulator may have relied on the more traditional, familiar display: the onboard weather representation. This discrepancy tends to obscure the trust findings somewhat, and suggests that each display may have unique advantages and disadvantages in the minds of individual flight crew members.

Perceived Workload. Workload effects were in agreement with our expectations, showing that perceived workload tended to be higher when the aircraft was closer to the weather event (see Figure 7). It was a bit of a surprise to see that perceived workload did not covary with weather display system agreement, as expected. One possible explanation for this finding is that participants considered the construct of workload to be more relevant to the immediate flight task itself than to the interpretation of the weather displays.

In truth, the magnitude of the ratings for workload was fairly low across the board. This is fairly intuitive; for the majority of the flights, participants were reliant on the autopilot to do the actual flying. The pilot flying took manual control of the aircraft only when weather events were presented. Even then, actual deviations were not required; rather, pilots were expected simply to maintain the flight path. Although it was heartening to observe that there was agreement between the NASA-TLX questionnaire, administered during each weather event, and our own single-item instrument, administered retrospectively, such agreement may actually help explain why we did not find differences in workload as a function of display agreement. Perhaps pilots were relying on retrospective memory to complete both workload questionnaires. If so, the fallibility of human memory may have led them to underestimate the workload associated with low weather display system agreement.

Perceived Situation Awareness. Perceived situation awareness varied somewhat with distance as did workload, suggesting that distance to weather was an important variable from a variety of perspectives. The noted interaction between weather distance and display agreement for situation awareness contributes to the notion that participants used the NEXRAD system to help them build a mental model of the outside world. The NEXRAD system, by its very nature, may have more to offer with regard to situation awareness than the onboard system. NEXRAD images are comprehensive and far-reaching; they depict weather cells in their entirety, along with the surrounding conditions. In contrast, the onboard depictions of weather are somewhat limited in scope. Onboard imagery allows flight crews to observe weather conditions along the immediate flight path; however, at distant ranges, it is not possible to see the full extent of weather cells. The problem is intensified at close ranges because the weather cell would take up practically the entire display, leaving the flight crew uncertain about surrounding conditions.

In actual flight situations, it is likely that flight crews would rely on additional information from air traffic control operators to determine situational status. This helps to explain the fairly low apparent variability in situation awareness scores in Figure 8. One interesting observation was that first officers seemed to retain greater situation awareness than captains. A number of explanations are possible for this finding, including the notion that first officers were more current with their training, or that they were more vigilant because they were in the presence of captains. It is difficult to resolve this question, however, without further information.

Unlike the workload questionnaire items, we did not observe consensus between the SART form and our one-item retrospective index of situation awareness on the opinion questionnaire. Although disappointing, this is not terribly surprising. Because the construct of situation awareness is considerably complex, it is likely that attempting to tap it by asking a simple question may not have been successful or reflective of the construct. To clarify, our single situation awareness item asked whether participants had complete knowledge of the flight environment. This implies that what was most important was the outside world, as depicted by the flight displays and the onboard and NEXRAD weather displays. In contrast, the SART asks respondents about a variety of aspects of situation awareness: attention level; arousal; situation instability, complexity and variability; and information quality and quantity, to name a few.

Weather Confidence. One variable that seems to supplement measures of display trust and measures of situation awareness is the notion of weather confidence. In this research, we asked participants to indicate the degree to which they believed that a displayed (or non-displayed) weather event actually existed. It is interesting that confidence level in the weather was greater when the captain was flying the aircraft. This may suggest a global trust effect associated with expert power (French & Raven, 1959). Pilot ratings also seemed to place more confidence in the onboard weather display than the NEXRAD system. Although this finding is counter to the finding for situation awareness, it agrees with comments made by the participants during the experimental sessions. It also tends to agree with the findings for trust in the displays – but only for the data for the pilot not flying (see Figure 6). As expected, participants seemed to have the greatest confidence in the existence of weather when both displays were in agreement. This is predictable given existing research that shows the superiority of redundant displays (Selcon, Taylor, & Shadrake, 1991).

The influence of weather distance on weather confidence was intriguing. As one might expect, confidence tended to rise as successive weather presentations were occurred and distance to the weather decreased. However, confidence seemed to be lowest when the first officer was flying and the range to weather was either very great (160 nm) or very small (20 nm). Because captains were the ultimate authority for decisions such as this one, it may be that they felt more confident about the collective decision if they were at the simulator controls, particularly in ambiguous circumstances (maximal or minimal distance).

Deviation Decision Confidence. This variable also indirectly reflects crews' trust of the weather displays. The findings here are a bit more striking than those for other variables. Crews showed greater confidence when the onboard weather system depicted

impending weather; regardless of whether or not the NEXRAD system showed similar information. In contrast, crews' confidence in the NEXRAD system was quite low when it alone showed weather information. An interesting aspect of these findings was that confidence levels appeared fairly low for both types of imagery systems when weather cells were 160 nm away. However, confidence appeared to spike for the onboard and the combination of onboard and NEXRAD systems at the 80 nm range. Confidence in the NEXRAD system, however, remained quite low at all weather ranges.

As might be expected, these findings converge with the findings for confidence that the weather event actually existed. Anecdotal comments made by the flight crews suggests that in practice they are most apt to heed onboard sources of weather, because the sensors driving such displays are on the aircraft itself, and are (presumably) better estimators of impending weather conditions. There is a considerable amount of research in the automation field to suggest that task operators are more likely to trust automated system actions if they understand the reasons why they occurred (Parasuraman & Riley, 1997; Lee and See, 2004). In the current experiment, the onboard weather system likely reflects the system that most aviators know. Therefore, because they are more comfortable with it, they are more likely to trust what it depicts, and follow directives warranted by it. They may also be skeptical of NEXRAD imagery because of the possibility for outdated databases or areas of poor terrain resolution as discussed by some researchers (Williams, Yost, Holland, & Tyler, 2002).

Communication. Of the measures of performance generated in this experiment, communication seemed to be the most equivocal. Yet, it is intriguing that the variables that best predicted communication level were generally associated with the captain: leadership style, captain age, captain experience (flight hours), and captain situation awareness. Much of this is likely due to the instructions we gave flight crews. To avoid ambiguities regarding command structure, we stipulated that the captain should be the final authority on all deviation decisions made. This undoubtedly forced the captain to take an active role to communicate.

Leadership Style. Existing research concerning leadership style suggests that teamed operators may benefit from using a participative leadership structure, because it allows the accuracy of the decision making to rise (see Bliss & Fallon, 2003). As described earlier, the majority of participants had never flown with each other prior to participating in the experiment (see Table 2). This was intentional, so that leadership style and communication level would more clearly reflect the particular influences of the flight task. Because they were unfamiliar with each other, it makes sense that they chose to adopt a participative leadership style most of the time. It also makes sense in terms of the task requirements. The results of the logistic regression may suggest that as captains age they are more likely to solicit input from younger first officers. Alternatively, it may be that older captains desired more interaction from first officers to make sure that the collective decisions were indeed democratic.

Anecdotal Observations. In general, we believe that the data we collected in this experiment were informative. Crews approached the experimental paradigm and tasks with an appropriate level of conscientiousness, and were forthcoming with their

reactions, and their suggestions and impressions of the flight scenarios used. As the experiment progressed and more pilot crews participated, we learned an increasing amount of information about weather confrontation and deviation scenarios in general.

The participants were eager to note that certain aspects of the experimental paradigm did not mirror an actual flight situation. For example, a clear omission was the role of air traffic control. Although it would have increased ecological validity to include a mock air traffic controller within the paradigm, we did not elect to do so. Had we included air traffic control directives, the variability associated with flight crew-controller dialogue would have likely eroded the internal validity of our manipulations, so that we would not have been able to cleanly measure the influences of display system agreement, weather range, and pilot flying.

It was also clear that our manipulation of weather range was artificial. Several pilots noted that in actual aircraft, they enjoy the freedom to select the range of the weather radar display as needed. Instead, we forced them to view the ranges that we prescribed, at the times we prescribed them. This setup allowed us to ensure that each crew was exposed to certain weather ranges for equal exposure times, and that no ranges were omitted. Yet, the artificiality of this technique probably detracted from the spontaneity of crew behaviors.

One of the most pronounced differences between the current experimental paradigm and an actual flight environment concerns the availability of preflight weather briefing material. As noted earlier, we provided flight crews with data portraying supposed weather conditions approximately 10 hours prior to their flight. However, the data were purposely vague, so as not to contaminate our manipulations of weather event presence. The pilots were graciously accommodating; however, it was clear that they were expecting more in the way of a pre-flight briefing. Perhaps in future research it would be best to attempt to match a preflight briefing with the situations that are to be depicted in flight. However, to effectively match such things would require considerable effort (in fact, may not be possible).

Obviously, actual weather encounters set in motion an extremely complex chain of interdependent, fluid events as flight crews select among alternatives for deviating from the weather path. Much of this complexity is due to the fact that weather scenarios are dynamic entities. Flight crew actions and aircraft movements do not happen in the sort of static vacuum that was represented in this experiment. However, with increased complexity comes unmanageable variability in pilot behavior, and unanticipated contingencies. We hope that our attempts to simplify a typically complex process have not made generalizations untenable. Indeed, we believe that pilot reactions as they have been observed in this experiment are likely reflective of actual flight situations. Not because of the veridicality of our experimental paradigm, but because of the professionalism of the experimental participants and the experimenters. We believe that participants exhibited the seriousness and thoughtfulness that are hallmarks of their day-to-day responsibilities. It is partly this professionalism that has caused researchers such as Gopher, Weil, Bareket, & Caspi (1988) to suggest that low or medium-fidelity flight simulation can successfully reflect actual flight behaviors.

One of the richest sources of information gathered during this experiment was the social interactions between members of the flight crews. As we described earlier, approximately half of the crews demonstrated a high level of communication as they

reacted to weather displays. Not coincidentally, crews also followed a participative leadership style the majority of the time. Such free exchange of ideas and directives likely stemmed from the fact that participants were eager to perform well, and from the recognition that each member had subject matter knowledge and a unique responsibility to perform that he could bring to bear on the flight task.

Contributions of this Research

Although there are numerous ways in which our experimental paradigm lacked fidelity, we believe that it offers much to theory in a number of areas, and to applied investigations of pilot reactions to weather.

Investigations of alarm and display trust have become more popular since the beginning of the 1990s. Almost without exception, the studies that have been conducted have featured naïve participants performing sterile research tasks. In recent years, however, researchers have attempted to use simulations of more realistic primary tasks such as medical monitoring and diagnosis (see Meyer, 2001), and process control (Muir, 1989). The current research marks the first time that alarm or display trust has been investigated by using actual aviators performing an aviation task. For that reason, and because of the compelling results reported here showing greater trust associated with display agreement, the current effort makes an important contribution to theoretical investigations of alarm and display trust.

Additionally, advances have been made here regarding the conceptual measurement of trust by questionnaires. Previously, experimenters have validated measures of trust in low-fidelity experimental paradigms using participants with limited task knowledge. For that reason, existing questionnaire measures lacked realism. For the current project, the research team adapted an existing questionnaire to more closely match the demands and the performance aspects of the aviation task. The results were quite encouraging. The trust measure appeared to be sensitive to manipulations of all independent variables, singly and in combination. There is also evidence to support its use in more general laboratory studies as well (Fallon, Bustamante, Ely, & Bliss, in press).

We are also enthusiastic about the potential contributions of this research to applied transport aviation. As technology continues to mature, cockpit displays of weather will undoubtedly become more complex and more visible in air transport cockpits. For flight crews to make the best use of these displays, it is important that designers understand their expectations and their tendencies to trust or mistrust; believe or question the information available on them. From the results reported here, several conclusions may be drawn to aid designers and users of cockpit weather displays. Because this project represents only a single data point, replication will be necessary. However, for now these statements are a reasonable starting point:

- Participants tended to place more confidence in the onboard weather depiction system, perhaps because it was the more familiar system. Interestingly, this confidence led them to make some incorrect deviation decisions in situations where the onboard did not display weather but the NEXRAD system did.

- Participants tended to respond to the weather displays in a way that signaled conservative decision making. They showed a tendency to deviate more often than not, and they were most likely to deviate when 40 nm from weather. Teams were more likely to make an accurate deviation decision when the NEXRAD system depicted impending weather.
- Distance to the weather events affected trust in the displays, though the observed relationships varied with the pilot flying, and the level of display agreement.
- Flight crew members experienced greater workload as they got closer to weather events.
- First officers reported more situation awareness than captains.
- Flight crews relied on a participative leadership style often, particularly if the captain was older.
- Communication level between flight crew members was most predictable from the captain's age and level of situation awareness.

Future Research Needs

The research reported here represents an initial examination of flight crew reactions to integrated weather displays. The results of our investigation were considerably complex; deviation decision accuracy, trust, workload, and situation awareness appeared to be determined by a multitude of factors. Although such findings likely reflect the true nature of weather reactions, it is clear that more investigation is needed to fully explore reactions to weather. In addition, we explored many constructs by using logistic regression. This tool may allow for prediction, but falls short of the goal of explanation. It is necessary to experimentally manipulate factors such as leadership style, communication level, workload, and situation awareness to ascertain their causative effects on weather display reactions.

It is also necessary to replicate the current findings in more complex flight situations. As discussed, we contrived our flight scenario to allow clean manipulation of weather distance, pilot flying, and weather display agreement. However, certain influences were not represented.

One of these influences is the presence of air traffic control. From conversations with the subject matter expert aviators and the participant flight crews, it is clear that deviation decisions are determined not just from appraisals of the weather displays, but by taking into consideration surrounding traffic, weather, and flight constraints voiced by air traffic control. It is not uncommon for flight crews to request a deviation, only to be countermanded by air traffic controllers. Such differences can lead to contention, and in some cases may affect the level of trust flight crews place on cockpit displays (Arri, 1991).

Air traffic control representation is one way to increase simulation realism. Another is to allow participants the freedom to change the range of the weather display at any time they choose. Although such freedom would likely preclude the manipulation of

range to weather as an independent variable, researchers may still be able to investigate its effect statistically, by covarying dwell time on particular ranges with deviation decision performance or with other performance aspects.

Perhaps the most complex but realistic change would be to allow flight crews to actually execute their deviation decisions. In the current experiment, we did not allow this because the resulting flight paths would not be comparable across flight crews. Yet, it became clear that there is a difference between deviation decision making and deviation decision implementation. Each of the three common decision rationales (safety, comfort and economy) are likely to impact the particular path crews choose around weather, how long the aircraft remains on that path, and the nature of the recovery from deviation. In short, allowing choice implementation would enhance the richness of our investigation, and would further clarify the variability surrounding flight crew trust.

In this experiment, we created onboard and NEXRAD weather depictions in an artificial manner, simply placing them at points along the flight path that seemed logical. However, conversations with flight crews suggested that the particular placement of weather cells may have led to some confusion in display interpretation. For example, several aviators noted that NEXRAD and onboard systems may not reliably depict weather when the terrain below the aircraft is mountainous, or when the flight path travels across vast expanses of water. This problem has been noted by some researchers in evaluations of general aviation weather displays (Williams, Yost, Holland, & Tyler, 2002). Such circumstances may represent sources of disagreement between onboard and NEXRAD weather representations, and therefore may become additional variables to consider in subsequent investigations of display trust.

REFERENCES

- Arri, E. (1991). Between a rock and a hard place: Weather deviations. ASRS Directline, 2. ONLINE: http://asrs.arc.nasa.gov/directline_issues/dl2_rock.htm.
- Beringer, D.B., & Ball, J.D. (2004). *The effects of NEXRAD graphical data resolution and direct weather viewing on pilots' judgments of weather severity and their willingness to continue a flight*. (Final Report No. DOT/FAA/AM-04/5). Oklahoma City, OK: Federal Aviation Administration.
- Beringer, D.B. & Schvanedeldt, R. (2002). Priorities of weather information in various phases of flight. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 6-10.
- Bliss, J.P. (1993). The cry-wolf phenomenon and its effect on alarm responses.
- Bliss, J.P. (2003). Collective mistrust of alarms. *International Journal of Applied Aviation Studies*, 3(1), 13-38.
- Bliss, J.P., Deaton, J.J., & Gilson, R.D. (1995). Alarm response behavior under conditions of varying alarm reliability. *Ergonomics*, 38(11), 2300-2312.
- Bliss, J.P., & Fallon, C. (2003). The effects of leadership style and primary task workload on team performance and follower satisfaction. *International Journal of Applied Aviation Studies*, 3(2), 259-276.
- Bliss, J.P., Jeans, S., & Prioux, H. (1996). Dual-task performance as a function of individual alarm validity and alarm system reliability information, *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Boyer, B., Campbell, M., May, P., Merwin, D., & Wickens, C.D. (1995). Three dimensional displays for terrain and weather awareness in the National Airspace System. *Proceedings of the 39th Annual Human Factors and Ergonomics Society Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Capobianco, G., & Lee, M. (2001). The role of weather in general aviation accidents: An analysis of causes, contributing factors and issues. *Proceedings of the 2001 Human Factors and Ergonomics Society Meeting*, October 8-12, Minneapolis, MN.
- Chemers, M. M. (2000). Leadership research and theory: A functional integration. *Group Dynamics: Theory, Research and Practice*, 4, 27-43.
- Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33(2), 205-31.
- Endsley, M. (1990). A methodology for the objective measurement of pilot situation awareness. In *Situation Awareness in Aerospace Operations (AGARD-CP-478)*. Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.
- Fallon, C.K., Bustamante, E.A., Ely, K., & Bliss, J.P. (in press). Improving user trust with a likelihood alarm display. *Proceedings of the 1st Annual Augmented Cognition International Conference*. July 22-27, 2005: Las Vegas, NV.

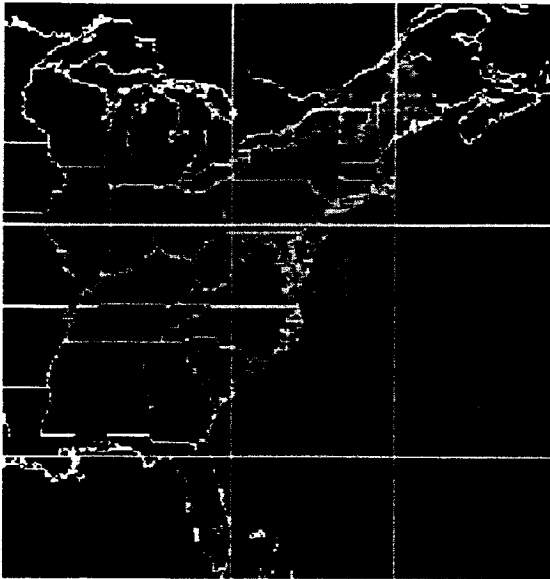
- Foushee, H.C. (1982). The role of communications, socio-psychological, and personality factors in the maintenance of crew coordination. *Aviation, Space, and Environmental Medicine*, November, 1062-1066.
- Foushee, C. H. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist*, 39, 885-893.
- French, J. R. & Raven, B. (1959). The basis of social power. In D. Cartwright, *Studies in Social Power*. Ann Arbor, University of Michigan.
- Getty, D.J., Swets, J.A., & Pickett, R.M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied*, 1, 19-33.
- Goh, J., & Wiegmann, D. (2001). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of the 2001 Human Factors and Ergonomics Society Meeting*, October 8-12, Minneapolis, MN.
- Gopher, D., Weil, M., Bareket, T., & Caspi, S. (1988). Fidelity of task structure as a guiding principle in the development of skill trainers based upon complex computer games. *Proceedings of the 1988 Human Factors Society Meeting*, Santa Monica, CA, 1266-1270.
- Gupta, N., Bisantz, A.M., & Singh, T. (2002). The effects of adverse condition warning system characteristics on driver performance: An investigation of alarm signal type and threshold level. *Behavior and Information Technology*, 21(4), 235-238.
- Hart, S., & Wickens, C.D. (1990). Workload assessment and prediction. In H.R. Booher (Ed.), *Manprint: An integrated approach to systems integration* (pp. 257-296). New York: Van Nostrand.
- Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology*, 9, 19-32.
- Jacob, R.J.K., Egeth, H.E., & Bevon, W. (1976). The face as a data display. *Human Factors*, 18, 189-200.
- Kantowitz, B.H., & Casper, (1988). Human workload in aviation. In E.L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 157-187). San Diego, CA: Academic Press.
- Latorella, K.A., & Chamberlain, J.P. (2002). Tactical vs. strategic behavior: General aviation piloting in convective weather scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 101-105.
- Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human Computer Studies*, 40, 153-184.
- Lee, J. D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- Liang, K. Y. & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 13-22.
- Lindholm, J.M. (1999). Weather information presentation. In D.J. Garland, J.A. Wise, & V.D. Hopkin (Eds.), *Handbook of aviation human factors*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., Jr., & Wherry, R.J. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies*. (Technical Report 851). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.
- McDonald, D.P., Gilson, R.D., & Mouloua, M. (1996). Spatial proximity of multiple alarms and the cry-wolf phenomenon. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43(4), 563-572.
- Muir, B.M. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. Doctoral thesis, University of Toronto, Canada.
- Muir, B.M. (1994). Trust in automation: Pt I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37 (11), 1905-1922.
- Nicholas, J. M., & Penwell, L. W. (1995). Proposed profile of the effective leader in human spaceflight based on findings from analog environments. *Aviation, Space, and Environmental Medicine*. 66, 63-72.
- O'Brien, J.V. & Wickens, C.D. (1997). Free flight cockpit displays of traffic and weather: Effects of dimensionality and data base integration. *Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Orasanu, J., Davison, J. & Fischer, U. (2001). The role of risk in aviation decision making: How pilots perceive and manage flight risks. *Proceedings of the 45th Annual meeting of the Human Factors and Ergonomics Society*. Santa Monica: Human Factors and Ergonomics Society.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230-253.
- Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.
- Roske-Hofstrand, R.J., & Paap, K.R. (1986). Cognitive networks as a guide to menu organization: An application in the automated cockpit. *Ergonomics*, 29, 1301-1311.
- Scanlon, C.H. (1992). *A graphical weather system design for NASA transport systems research vehicle B-737*. (NASA-TM-104205). Hampton, VA: NASA Langley Research Center.
- Selcon, S.J., Taylor, R.M., & Shadrake, R.A. (1991). Giving the pilot two sources of information: help or hindrance? In E. Farmer (Ed.), *Human Resource Management in Aviation*, p. 139-148. Aldershot: Avebury Technical.
- Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: The links between language, performance, error, and workload. *Journal of Human Performance in Extreme Environments*. 5, 63-68.
- Sherman, P.J. (2003). Applying crew resource management theory and methods to the operational environment. In P.S. Tsang & M.A. Vidulich, *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sly, J., & Hartmann, G. (1999). *Weather avoidance using route optimization as a decision aid: An AWIN topical study*. Phase I Final Report. Honeywell Technology Center. Minneapolis, MN.

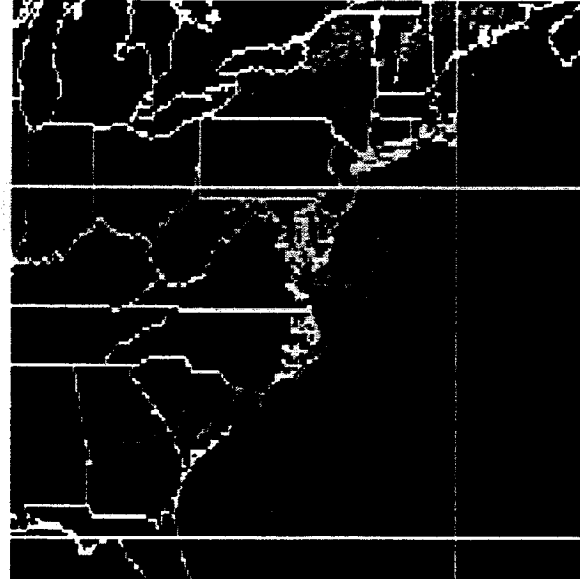
- Stokes, A., & Wickens, C.D. (1988). Aviation displays. In E.L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 157-187). San Diego, CA: Academic Press.
- Tabachnick, B. & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed). New York: HarperCollins.
- Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (AGARD-CP-478)(pp.3/1-3/17). Neuilly Sur Seine, France: NATO-AGARD.
- Tsang, P. & Vidulich, M.A. (2003). *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vidulich, M.A. (2003). Mental workload and situation awareness: Essential concepts for aviation psychology practice. In P.S. Tsang & M.A. Vidulich, *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vroom, V. H., (2000). Leadership and the decision-making process. *Organizational Dynamics*, 28, 82-94.
- Vroom, V. H., Jago, A. G. (1978). On the validity of the vroom-yetton model. *Journal of Applied Psychology*. 63, 151-162.
- Wickens, C.D. (1995). Aerospace techniques. In Weimer. J. (Ed.), *Research Techniques in Human Engineering*. (pp. 114- 118). Upper Saddle River, N.J.: Prentice Hall PTR.
- Wickens, C.D. (2000). Human factors in vector map design: The importance of task display dependence. *Journal of Navigation*, 53, 1-14.
- Wickens, C.D. (2003). Aviation displays. In P.S. Tsang & M.A. Vidulich, *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wickens, C.D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wickens, C.D., Kroft, P., & Yeh, M. (2000). Database overlay in electronic map design. In *Proceedings of the IEA 2000/HFES 2000 Congress*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Wiggins, M. (1999). The development of computer-assisted learning (CAL) systems for general aviation. In D. O'Hare (Ed.), *Human performance in general aviation* (pp. 153-172). Aldershot, England: Ashgate.
- Williams, K.W., Yost, A., Holland, J., & Tyler, R.R. (2002). *Assessment of advanced cockpit displays for general aviation aircraft – the Capstone Program*. (Final Report No. DOT/FAA/AM-02/21). Washington, D.C.: Federal Aviation Administration.

APPENDIX 1

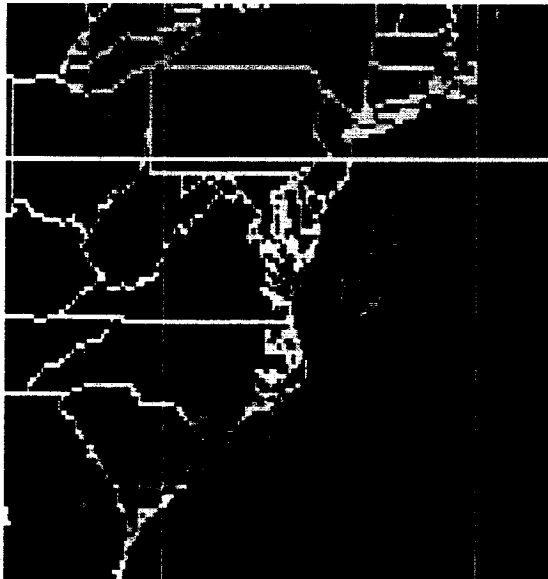
SAMPLE ONLINE NEXRAD WEATHER PRESENTATIONS



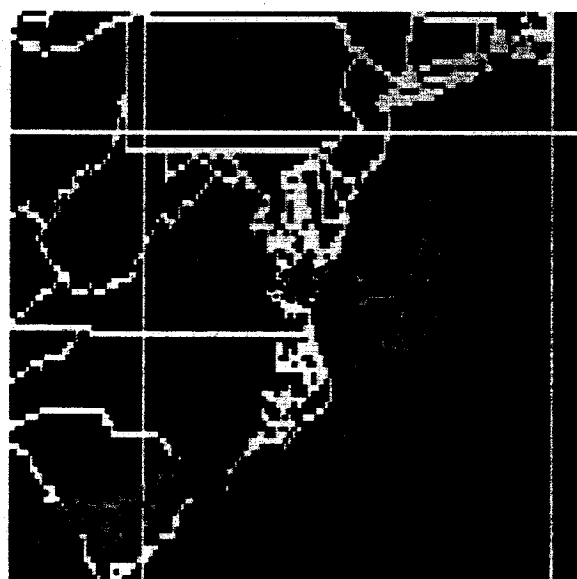
160-Mile Depiction of Weather



80-Mile Depiction of Weather



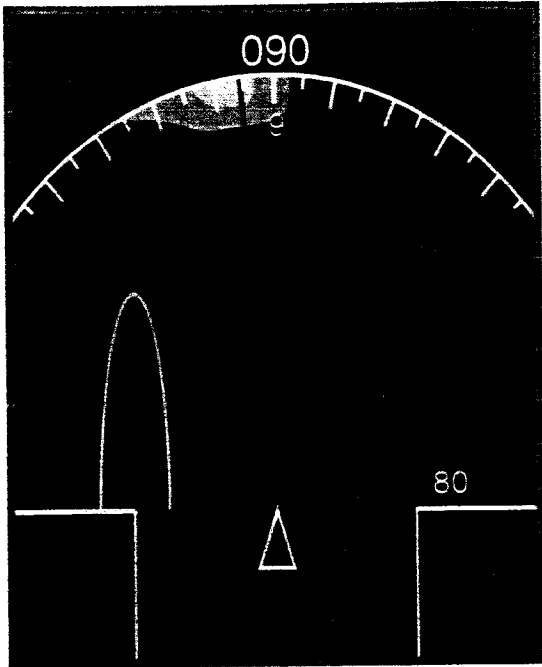
40-Mile Depiction of Weather



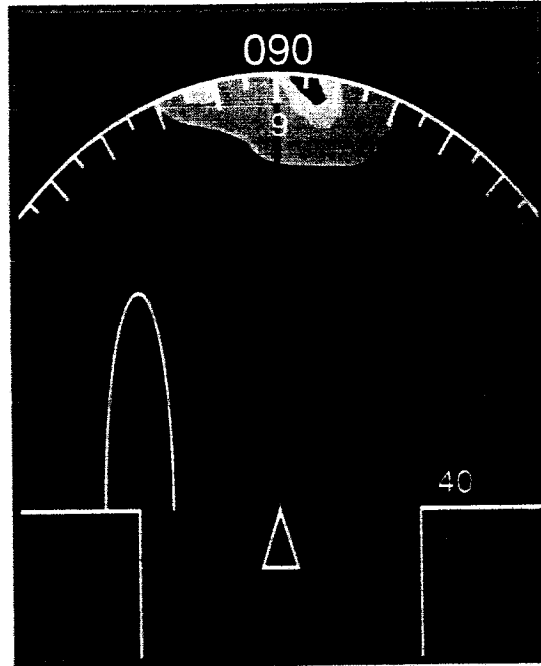
20-Mile Depiction of Weather

APPENDIX 2

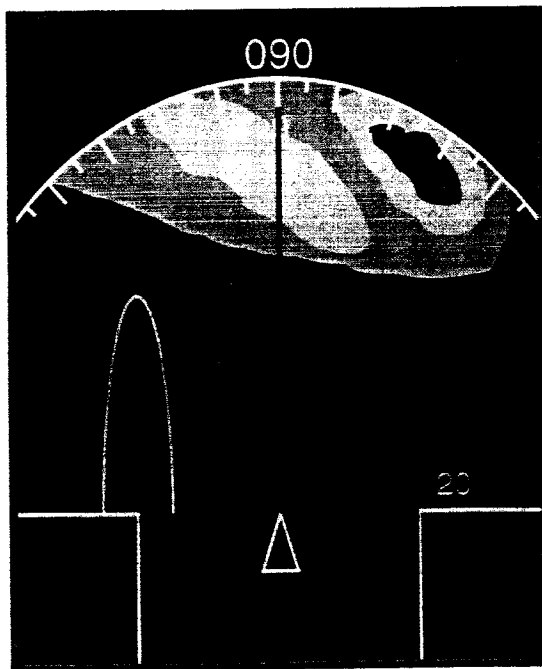
SAMPLE ONLINE ONBOARD WEATHER PRESENTATIONS



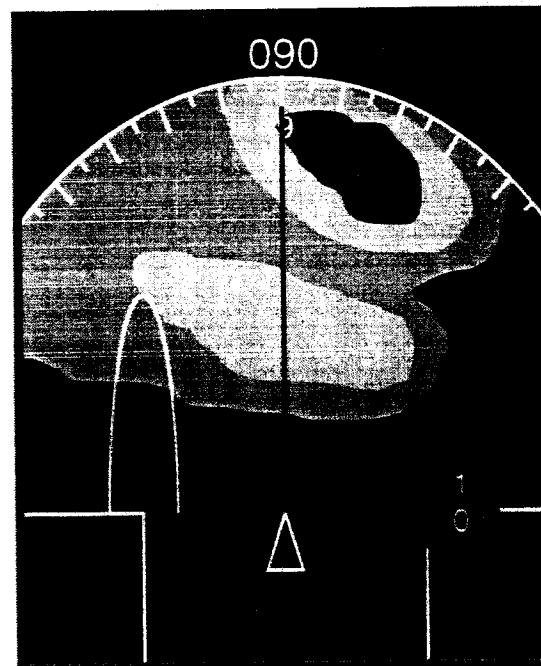
160- Mile Depiction of Weather



80-Mile Depiction of Weather

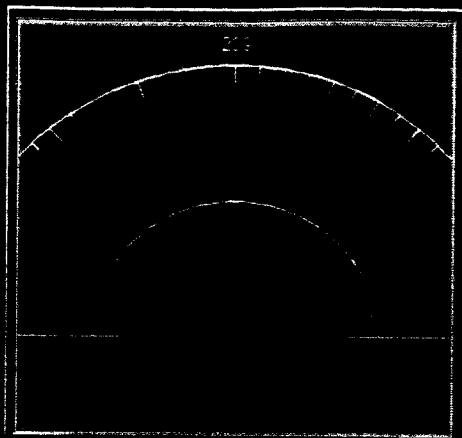


40-Mile Depiction of Weather

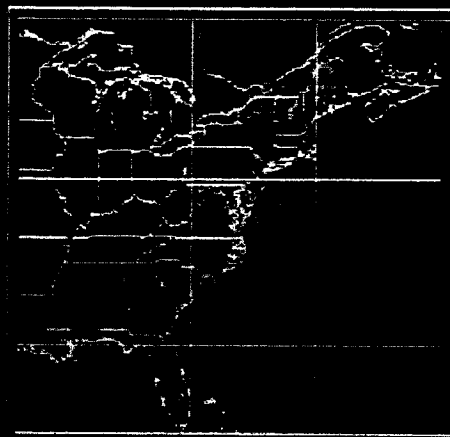


20-Mile Depiction of Weather

ONBOARD



NEXRAD



0° 100°

0° 100°

• Not Centered

• Elevation Low

• Deviate Right

0° 100°



APPENDIX 3

ONLINE SART SITUATION AWARENESS QUESTIONNAIRE

Instructions

Situation Awareness refers to your ability to relate the meaning of events and elements in an uncertain environment to mission goals and objectives. The technique involves the scoring of ten different scales, each of which is potentially a factor in your Situation Awareness.

Remember the scales are a subjective measure of your individual perceptions during the simulated flight in the context of your experience with flying in general. There is no right or wrong answer to give, only your best estimate of your personal experience from the point of view as a pilot. Do not spend too much time on any one item. Your initial 'gut feeling' is likely to be the most accurate estimation.

The following are the definitions of each of the 10 SART rating items. Please read through these until you are sure you understand their meanings. Feel free to ask the experimenter if you are unsure of any of these definitions. Refer to these descriptions as you do the ratings.

Please **indicate the number** that best describes your level of situation awareness for each dimension.

1. Instability of Situations (D)

To what extent were the situations and environmental factors encountered through the course of the flight likely to change? Were they very dynamic and likely to change suddenly (High), or were most of them slow and stable with easily predictable outcomes (Low)?

2. Complexity of Situations (D)

How complicated were events during the flight? Were they complex with many closely interrelated components and/or phases (High), or were most simple and straight forward with few interrelated components and/or phases (Low)?

3. Variability of Situations (D)

On average, how many elements were changing at any one time? Were there a large number of dynamic variables (high), or very few that might change at once (low)?

4. Arousal (S)

How alert and ready for action did you feel throughout the course of the exercise? Could you anticipate the flow of events and respond quickly (high), or were you hard pressed to keep up with evolving situations (low)?

5. Spare Mental Capacity (S)

How much mental capacity did you have to spare in this flight? Do you think you could have dealt with a significant number of additional elements and variables if necessary (High), or did the complexity of the flight take all your mental capacity combined with available decision aids and analysis tools to handle (low)?

6. Concentration of Attention (S)

How much could you concentrate your attention in each problem situation? Were your thoughts always focused on important elements and events (high), or did internal and external factors distract you and draw your attention elsewhere (low)?

7. Division of Attention (S)

Were you able to divide your attention among several key issues during the course of the flight? Were you usually concerned with many aspects of current and future events simultaneously (high), or did you focus on only one thing at a time (low)?

8. Information Quantity (U)

How much useful information were you able to obtain from all available sources during the flight? Did you receive and understand a great deal of pertinent data (high), or did you receive and understand very little (low)?

9. Information Quality (U)

How good was the information you obtained about the situation? Was the communicated knowledge very valuable (high), or was the communicated knowledge not helpful (low)?

10. Familiarity with Environment (U)

How familiar were you with the different elements and events in the environment and situations encountered during the course of this flight? Could you call on a great deal of relevant experience and knowledge to fill in gaps in the available information (high), or did you find many aspects of the exercise new and unfamiliar to you (low)?

Overall Situation Awareness

Evaluate your awareness of the overall meaning of events and elements in the environment to the mission plan and eventual accomplishment of mission goals. Did you always have a complete picture and a plan for how the various elements would affect the

APPENDIX 4

ONLINE NASA-TLX WORKLOAD QUESTIONNAIRE

NASA TLX Rating Instructions

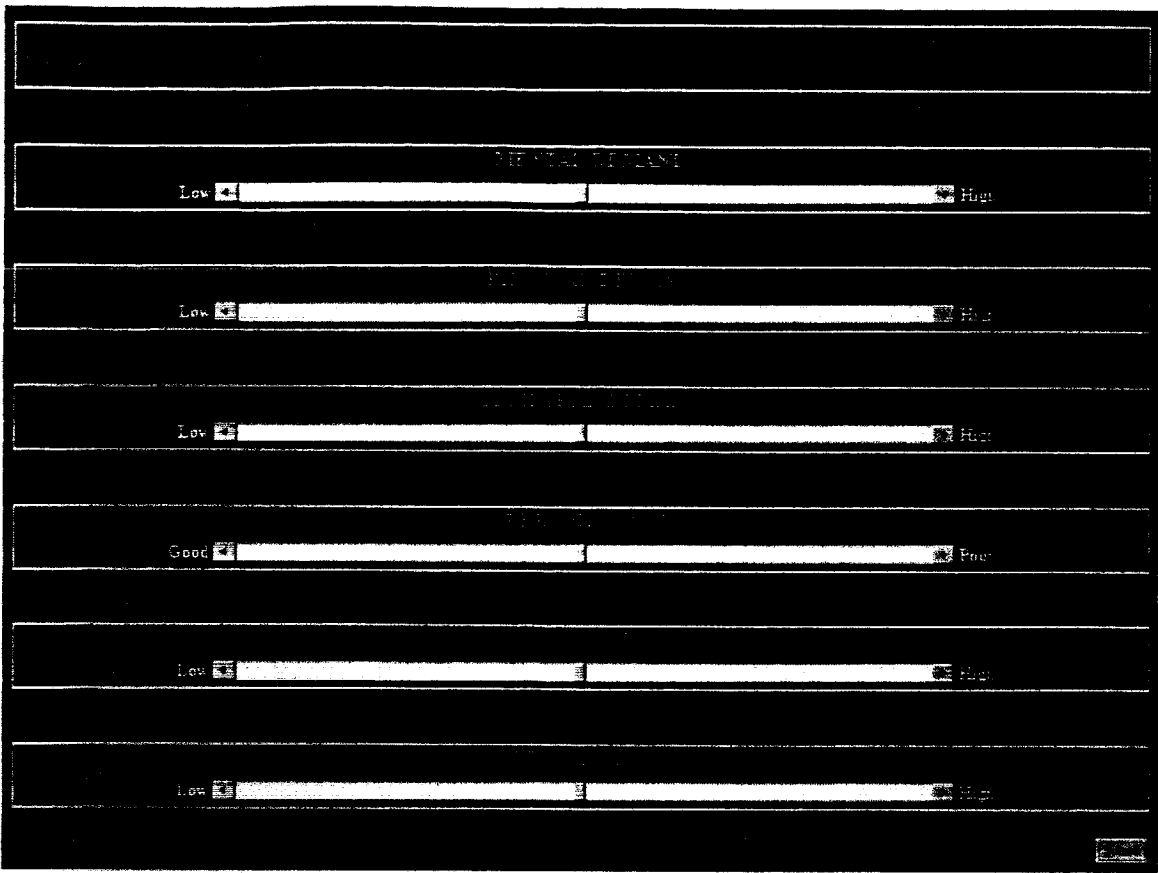
We are not only interested in assessing your performance but also your experiences during the different task conditions. In the most general sense, we want to examine the “workload” you experience. Workload is a difficult concept to define precisely, but a simple workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The workload contributed by different task elements may change. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are no effective “rulers” that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. A set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please take a moment to read the descriptions of the scales carefully (*see the back of this sheet*).

If you have any questions about any of the scales in the table, please ask the experimenter about them. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

After performing each of the tasks, you will be presented with a screen containing a set of rating scales. You will evaluate the task by placing an arrow on each of the six scales at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that “own performance” goes from “good” on the left to “bad” on the right. This order has been confusing for some people. Please consider your responses carefully in distinguishing among the different task conditions. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment and is greatly appreciated by all of us.

If you have any questions, please ask them now. Otherwise, start whenever you are ready. Thank you for your participation.



APPENDIX 5

ONLINE TRUST QUESTIONNAIRE

Below is a list of words used to describe trust in the onboard and ground weather information presented during the flight leg you have just completed. These words will also appear on your computer monitor along with a rating scale beneath each word. Please rate the words on extent to which you believe they describe the weather information. Use the mouse on your workstation computer to click the appropriate point on each scale. You will be asked to complete this questionnaire twice, once for the onboard display and once for the NEXRAD display. Remember, your ratings should only reflect your experience with the displays during the most recent weather presentation. The definitions for each word have been provided below. You may refer to these definitions to help you with your ratings.

Inconsistent – the system's behavior is erratic

Unpredictable – the system's future behavior is unknown

Truthful – the information presented by the system corresponds to reality

Accurate – the system performs without error

Trustworthy – the system's behavior is reliable

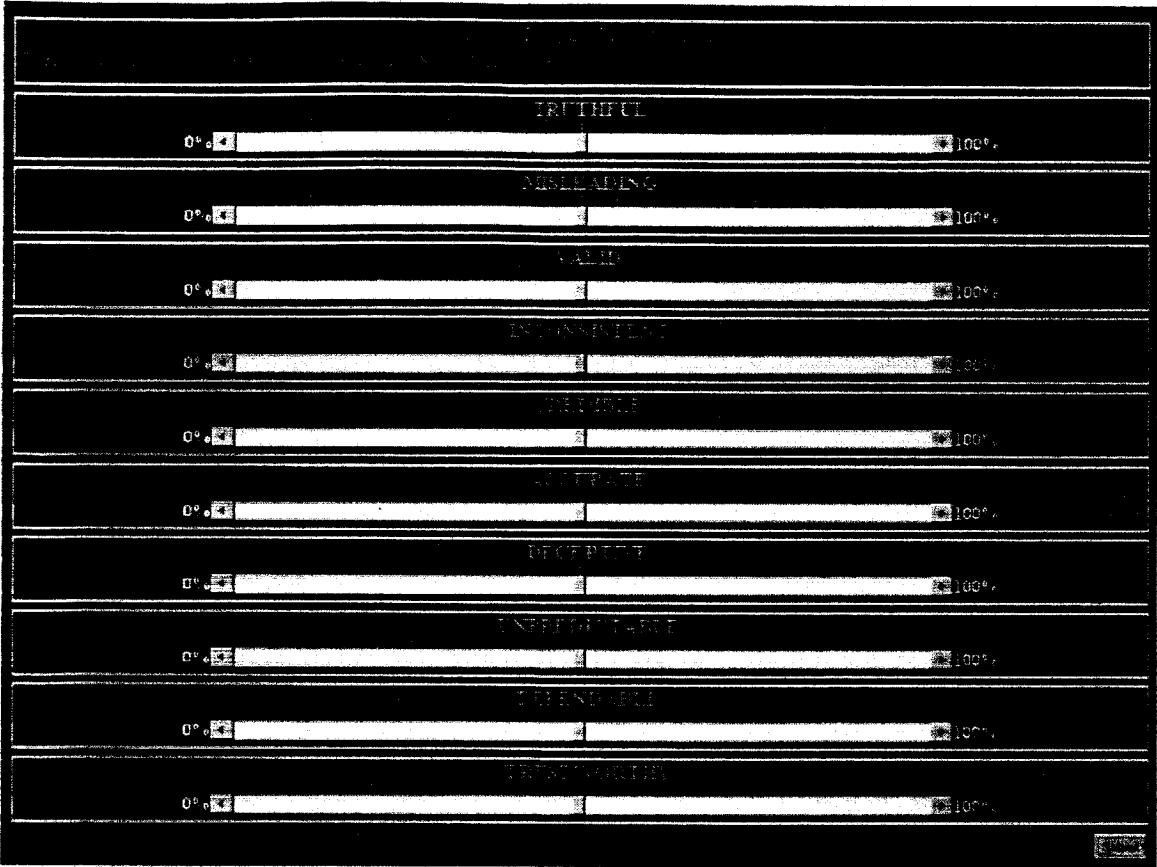
Misleading – the system leads one to commit errors

Deceptive – the system causes one to believe what is not true

Credible – the system is worthy user confidence

Valid – the correct actions can be inferred from the information presented by the system

Dependable – the system is worthy of user trust



APPENDIX 6

PRE-FLIGHT BRIEFING INFORMATION

WEATHER BRIEFING INFORMATION - TRAINING FLIGHT

Flight Specific Weather Package

Flight # 0001, SMF - LAX

Alternates

T/O: NONE

Landing: NONE

Driftdown: NONE

Arrival Information: N/A

Alternate Information: N/A

Hazard Information: NONE

Enroute Information: There have been scattered reports of convection enroute; however, complete details have not been provided.

Departure Information: N/A

Additional Information: This weather briefing information is approximately 10 hours old.

Flight Specific Weather Package

Flight # 0001, LAX - SMF

Alternates

T/O: NONE

Landing: NONE

Driftdown: NONE

Arrival Information: N/A

Alternate Information: N/A

Hazard Information: NONE

Enroute Information: There have been scattered reports of convection enroute; however, complete details have not been provided.

Departure Information: N/A

Additional Information: This weather briefing information is approximately 10 hours old.

WEATHER BRIEFING INFORMATION - EXPERIMENTAL FLIGHT

Flight Specific Weather Package

Flight # 0001, JFK - MIA

Alternates

T/O: NONE

Landing: NONE
Driftdown: NONE

Arrival Information: N/A
Alternate Information: N/A
Hazard Information: NONE
Enroute Information: There have been scattered reports of convection enroute; however, complete details have not been provided.
Departure Information: N/A
Additional Information: This weather briefing information is approximately 10 hours old.

Flight Specific Weather Package
Flight # 0001, MIA – JFK
Alternates
T/O: NONE
Landing: NONE
Driftdown: NONE

Arrival Information: N/A
Alternate Information: N/A
Hazard Information: NONE
Enroute Information: There have been scattered reports of convection enroute; however, complete details have not been provided.
Departure Information: N/A
Additional Information: This weather briefing information is approximately 10 hours old.

APPENDIX 7

PAPER AND PENCIL DEMOGRAPHIC BACKGROUND QUESTIONNAIRE

Part. #: _____ Group: _____ Team: _____ Date: _____ Time: _____

The purpose of this questionnaire is to collect background information for participants in this experiment. This information will be used strictly for this experiment and for research purposes only. Please complete each item to the best of your ability.

1. Age _____
2. Sex _____ (0=Male, 1=Female)
3. Have you ever been diagnosed as color blind or deficient? _____ (0 = No, 1 = Yes)
4. Have you ever been diagnosed as having hearing loss? _____ (0 = No, 1 = Yes)
5. How many hours per week do you use computers (work and recreation combined)?

6. How many hours/week do you play video/simulation games? _____
7. About how many total flight hours have you logged (including all types of aircraft)?

8. About how many flight hours have you logged in glass cockpit aircraft? _____
9. Please circle the types of aircraft ratings that you currently hold:
 1. Private
 2. Instrument
 3. Multi-engine
 4. Commercial
 5. Rotary
 6. CFI
 7. CFII
 8. Other _____
10. When was the date of your last FAA check ride? _____
11. What is your current rank? _____ (0=Copilot, 1=Captain, 2=other)
12. Have you ever interacted with an integrated weather display before? _____ If so, please list the specific system(s) you have encountered _____

13. How well do you know the other pilot? _____

1. I've never met him or her before today.
2. I barely know him or her (we've seen or met each other before, but have never flown together).
3. I know him or her fairly well (we've met and see each other occasionally, and have flown together 1-5 times).
4. I know him or her quite well (we see each other often and have flown together 5-10 times).
5. I know him or her extremely well (we know each other professionally and socially, and/or have flown together more than 10 times)

14. About how many times per week do you engage in teamed sporting activities (football, basketball, etc.)? _____

15. About how many times per week do you engage in teamed music activities (playing in bands, etc.)? _____

16. Do you do any other team-related activities? _____ If so, please describe them, and indicate how often you do them.

APPENDIX 8

PAPER AND PENCIL OPINION QUESTIONNAIRE

#: _____ Group: _____ Team: _____ Date: _____ Time: _____

Thank you for participating in this research project. Please complete the following items by entering the number of your choice on the answer sheet. Your answers are completely confidential.

Please rate the flight simulation on the following dimensions:

1. Visual Information

1. Very Realistic (all critical elements of the flight display were available and functioned predictably)
2. Slightly Realistic (some critical elements of the flight display were available, most functioned predictably)
3. Neither Realistic nor Artificial (there were some display elements present, but many were not)
4. Slightly Artificial (the flight display lacked many essential features and lacked functionality)
5. Very Artificial (the flight display was completely unrealistic and bore no relation to an actual flight task)

2. Auditory Information

1. Very Realistic (all critical elements of the flight display were available and functioned predictably)
2. Slightly Realistic (some critical elements of the flight display were available, most functioned predictably)
3. Neither Realistic nor Artificial (there were some display elements present, but many were not)
4. Slightly Artificial (the flight display lacked many essential features and lacked functionality)
5. Very Artificial (the flight display was completely unrealistic and bore no relation to an actual flight task)

3. Tactile/Motor Information

1. Very Realistic (controls operated similarly to an actual aircraft of the same type)
2. Slightly Realistic (controls operated realistically for the most part, but there were some actions that were artificial)
3. Neither Realistic nor Artificial (the most critical controls were functional, but most others were inoperable or artificial)
4. Slightly Artificial (the operation of controls was far removed from an actual flight experience)
5. Very Artificial (there's no comparison between the simulator controls and those in an actual aircraft)

4. Situation Awareness

1. Very Complete (I was able to understand the state of the flight environment totally)
2. Slightly Complete (I had adequate knowledge of most critical elements of the flight environment)
3. Neither Complete nor Incomplete (my knowledge of the flight environment was complete in some areas, but not others)
4. Slightly Incomplete (there were significant gaps in my knowledge of the flight environment)
5. Very Incomplete (my knowledge of the flight environment was unacceptably poor)

5. Experienced Workload Across the Entire Flight

1. Very High (I felt as if I had too much to do throughout the entire flight)

2. Slightly High (there were several periods when I felt overburdened)
3. Neither High nor Low (occasionally I felt overburdened by the flight, but I was able to compensate)
4. Slightly Low (at times the demands of the flight were excessive)
5. Very Low (the requirements of the flight were unreasonable at all times)

Please rate the weather presentation system on the following dimensions:

6. Realism:

1. Very Realistic (was exactly like other weather presentation systems I have encountered)
2. Slightly Realistic (resembled some other systems I've seen, but there were minor differences)
3. Neither Realistic nor Artificial (although the system resembled other systems, there were also important differences)
4. Slightly Artificial (there were some critical differences between this system and existing systems)
5. Very Artificial (this system bore no resemblance to any other weather display system I've seen before)

7. Comprehensiveness:

1. Very Comprehensive (all important weather elements were represented)
2. Slightly Comprehensive (most important weather elements were represented)
3. Neither Comprehensive Nor Limited (some important elements were represented, but some were missing)
4. Slightly Limited (most important weather elements were missing)
5. Very Limited (there were no important weather elements presented)

8. Disturbance:

1. Very Distracting (the presentation of weather messages was overly distracting)
2. Slightly Distracting (the incoming weather messages were inconvenient, but ultimately helpful)
3. Neither Distracting Nor Helpful (incoming weather messages made it somewhat difficult to concentrate on flying)
4. Slightly Helpful (incoming weather messages complemented my ability to fly)
5. Very Helpful (incoming weather messages made it easier for me to fly the airplane)

9. Reliability/Trustworthiness

1. Very Reliable (I trusted the incoming weather messages implicitly)
2. Slightly Reliable (I trusted most of the incoming weather messages, but some were not believable)
3. Neither Reliable nor Unreliable (I found myself believing about half of what I saw/heard)
4. Slightly Unreliable (it was difficult for me to take the weather messages seriously)
5. Very Unreliable (the system presented messages that were not believable)

10. Drive to respond

1. I did not feel compelled to modify my flight actions after the weather messages
2. I felt slightly compelled to modify my flight actions after the weather messages
3. I felt moderately compelled to modify my flight actions after the weather messages
4. I felt greatly compelled to modify my flight actions after the weather messages
5. It was imperative that I change my flight behavior following the weather messages

11. Did you have a strategy for reacting to the weather messages?

_____ (0=no, 1=yes)

If so, what was it? _____

12. Did you have any problems interacting with the other crewmember? _____

(0=no, 1=yes)

If so, please describe them _____

13. Do you have any other thoughts, feelings, or comments about this experiment?

APPENDIX 9

INFORMED CONSENT FORM

OLD DOMINION UNIVERSITY INFORMED CONSENT FORM

INFORMED CONSENT DOCUMENT

The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES.

TITLE OF RESEARCH: Pilot Trust of Weather Information

RESEARCHERS:

James P. Bliss, Ph.D., Associate Professor, College of Sciences, Psychology Department
Ernesto A. Bustamante, Graduate Student, College of Science, Psychology Department
Corey K. Fallon, Graduate Student, College of Science, Psychology Department

DESCRIPTION OF RESEARCH STUDY:

Display of information in the cockpit has long been a challenge for aircraft designers. Given the limited space in which to present information, designers have had to be extremely selective about the types and amount of flight related information to present to pilots. Important also is the timing of information display, and the integration of displayed information with existing information sources within the cockpit. Presenting even relevant information too soon may lead to complacency; presenting information too late may lead the pilot to miss critical signals or fail to react in time. The role of weather displays is to present near-real-time ("nowcasting") and predictive information about weather anomalies. In many cases, such presentation includes generating visual and auditory alarm signals to draw the flight crew's attention to potential weather-related problems. In this research, you will be required to fly four simulated routes while reacting to weather events presented on a separate visual display. Your flight performance will be measured, as will your reactions to the weather events. You will be videotaped when you are performing your simulated flight mission, to allow the experimenters to easily analyze your data. The results of the proposed research should allow NASA to make more informed decisions regarding the format and implementation of weather displays in cockpits, and should contribute to existing theories of alert reliability and display perception.

As part of this experiment, you will be asked to fill out a background information form, complete four sessions of a computer task, and answer a brief opinion questionnaire about the experiment. You will also be required to complete questionnaires regarding your trust in the weather information, cognitive workload and situation awareness. You

will complete the experiment with another pilot. The simulated flights (2 ½-hour and 2 2-hour flights) will last approximately six hours (three hours per leg); the entire experiment will last approximately 8 hours.

EXCLUSIONARY CRITERIA:

To participate, you must have normal vision or corrected-to-normal vision. You must also have normal or corrected-to-normal hearing. Therefore, if you normally wear eyeglasses, contact lenses or hearing aids you will need to wear them to participate.

RISKS AND BENEFITS:

RISKS: The risks from this study are similar to those associated with normal computer usage. However, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: If you decide to participate in this study, you will receive payment as agreed to through your arrangement with Lockheed Martin. You will also benefit by learning about weather display issues.

COSTS AND PAYMENTS:

As stipulated above, you will be compensated monetarily for participation in this project.

CONFIDENTIALITY:

Your participation in this research will be held confidential by the experimenter. Researchers will remove identifiers from the information. The results of this study may be used in reports, presentations, and publications; but researchers will not identify you. Additionally, individual results will not be made available to your employer. Videotapes will be erased immediately after the data have been coded and analyzed. Of course, your records may be subpoenaed by court order or inspected by government bodies with oversight authority.

WITHDRAWAL PRIVILEGE:

It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study -- at any time. Your decision will not affect your relationship with Old Dominion University, NASA Langley Research Center, or Lockheed Martin, or otherwise cause a loss of benefits to which you might otherwise be entitled. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation.

COMPENSATION FOR ILLNESS AND INJURY:

If you agree to participate, your consent in this document does not waive any of your legal rights. However, in the event of harm, injury, or illness arising from this study,

neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact James P. Bliss at 757-683-4222 or Dr. David Swain from the Old Dominion University Institutional Review Board, 757-683-6028.

VOLUNTARY CONSENT:

By agreeing to participate, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researcher should be able to answer them:

James P. Bliss at 757-683-4222

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. David Swain, at 757-683-6028, or the Old Dominion University Office of Research and Graduate Studies, at 757-683-3460.

By signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

Participant's Name Participant's Signature Date

INVESTIGATOR'S STATEMENT:

I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

Investigator's Name Investigator's Signature Date

APPENDIX 10

PARTICIPANT INSTRUCTIONS

Today you will be completing two simulated round trip flights – a practice flight from Sacramento, CA to Los Angeles, CA and back, and an experimental flight from New York, NY, to Miami, FL, and back. To complete this mission, you will be using this simulator, loaded with Microsoft Flight Simulator 2004. First of all, we need you to read this Informed Consent Form (see Attachment 1). If you have any questions, please do not hesitate to ask [*answer any questions that participants may have*]. If you agree to participate, we need you to complete this Background Information Form (see Attachment 2).

Before you begin your flight, you will go through a briefing session in which we will provide you with all the necessary information about the type of airplane you will be flying and the flight plan you need to follow. Next, you will go through a practice session, which will allow you to familiarize yourself with the flight simulator. Once you have completed this practice session, you will begin your mission, which will last approximately 2.5 hours each way for a total of five hours. After you finish the first part of the mission, you will receive a one-hour break. After the break, you will complete the second part of the mission. Throughout the mission, you will encounter potential weather events, which will indicate the presence of different upcoming weather events. These signals will be displayed on this computer [*point to it*]. However, the information is not entirely reliable. The data obtained to generate the information will vary with regard to how old it is. The data will be presented at different distances from the aircraft. Your job consists of indicating whether and how you would deviate, given the information. However, since the alarm system is not 100% reliable, **how much you rely on it to make your decisions is entirely up to you** [*emphasize this*]. You will also be required to complete some questionnaires after each presentation related to your level of workload, situation awareness, and trust in the weather information.

Throughout the flight, you are going to be videotaped. After you complete the flight, we will review the tape to help us determine more about your intended actions during the flight. Once you have completed the experimental flight, you will complete an opinion questionnaire regarding certain aspects of your performance and your interaction with the weather warning system (see Attachment 4). Lastly, you will go through a debriefing session, in which we will discuss the purpose of this study and answer any questions you may have.

APPENDIX 11

COLLABORATIVE DECISION MAKING RATING SYSTEM

Rater #:	Weather Event	Distance From Weather Event	Communication Level	Leadership Style
Team #:	1	160		
		80		
		40		
		20		
	2	160		
		80		
		40		
		20		
	3	160		
		80		
		40		
		20		
	1	160		
		80		
		40		
		20		
	2	160		
		80		
		40		
		20		
	3	160		
		80		
		40		
		20		

APPENDIX 12

CONSENSUS RATINGS OF COMMUNICATION AND LEADERSHIP STYLE

Rationale Communication Leadership Style
 1- safety 1- low 1- Participative
 2- comfort 2- high 2- Autocratic
 3- economy
 4- trust

Team	Event	Distance	Rationale	Communication	Leadership Style
3	1	160	4	2	1
		80	4	2	1
		40	4	1	2
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	3	2	1
		20	4	2	1
	3	160	4	2	1
		80	1	2	1
		40	4	2	1
		20	4	2	1
3	1	160	4	2	1
		80	4	1	1
		40	4	1	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
5	1	160	4	1	1
		80	4	2	1
		40	4	1	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1

	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
5	1	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
6	1	160	4	1	2
		80	4	1	2
		40	4	1	1
		20	4	1	2
	2	160	4	1	1
		80	4	1	2
		40	4	2	1
		20	4	1	2
	3	160	4	1	2
		80	4	1	2
		40	4	1	2
		20	4	1	2
6	1	160	4	1	1
		80	4	1	1
		40	1	1	2
		20	1	1	1
	2	160	4	1	2
		80	4	1	2
		40	1	1	2
		20	4	1	2
	3	160	1	1	1
		80	1	1	1
		40	1	1	1
		20	1	1	1
7	1	160	4	1	1
		80	4	2	1
		40	4	2	1

		20	4	2	1
	2	160	4	1	1
		80	4	2	1
		40	4	2	1
		20	4	2	2
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
7	1	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	1	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
8	1	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
8	1	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	2
	2	160	4	1	1
		80	4	2	2
		40	4	1	1
		20	4	1	2
	3	160	4	1	1
		80	4	1	1

		40	4	1	1
		20	4	1	1
9	1	160	4	1	2
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	3	2	1
		80	4	2	1
		40	4	1	1
		20	4	2	1
	3	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
9	1	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
	2	160	4	1	1
		80	4	1	1
		40	4	2	2
		20	4	1	1
	3	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
10	1	160	4	1	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	3	2	1
		80	4	2	2
		40	4	2	1
		20	4	1	1
	3	160	4	1	2
		80	4	1	2
		40	4	1	2
		20	4	1	2
10	1	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	1	1
	2	160	4	1	1

		80	4	2	1
		40	3	1	1
		20	3	1	1
	3	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
11	1	160	4	2	2
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	2	1
		20	4	1	2
11	1	160	4	1	1
		80	4	2	1
		40	4	2	1
		20	4	2	1
	2	160	1	2	1
		80	4	2	1
		40	4	2	1
		20	4	1	1
	3	160	4	2	1
		80	4	2	1
		40	4	1	1
		20	4	2	1
12	1	160	4	1	2
		80	4	1	1
		40	4	1	2
		20	4	1	1
	2	160	4	1	2
		80	4	1	2
		40	4	1	2
		20	4	1	2
	3	160	4	1	2
		80	4	1	2
		40	4	1	2
		20	4	1	2

12	1	160	4	1	1
		80	4	1	2
		40	4	1	1
		20	4	1	2
	2	160	4	1	2
		80	4	2	2
		40	4	1	2
		20	4	2	2
	3	160	4	2	1
		80	4	1	1
		40	4	2	1
		20	4	1	2
13	1	160	4	1	1
		80	4	1	1
		40	4	2	1
		20	4	2	1
	2	160	4	1	2
		80	4	1	1
		40	4	2	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	4	1	2
		20	4	1	2
13	1	160	4	1	1
		80	4	1	2
		40	4	1	2
		20	4	1	2
	2	160	4	2	1
		80	1	2	1
		40	4	2	2
		20	4	2	2
	3	160	4	1	2
		80	4	1	1
		40	4	1	2
		20	4	1	1
14	1	160	4	1	2
		80	4	1	1
		40	1	2	2
		20	4	1	1
	2	160	4	2	1
		80	4	2	2
		40	4	1	1

		20	4	2	1
	3	160	4	1	2
		80	4	1	1
		40	4	2	1
		20	4	2	2
14	1	160	1	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
	2	160	4	1	2
		80	4	1	2
		40	4	1	1
		20	4	1	2
	3	160	4	1	1
		80	4	1	1
		40	4	1	2
		20	4	1	2
15	1	160	4	1	1
		80	4	2	1
		40	4	2	1
		20	4	1	1
	2	160	1	2	1
		80	4	2	1
		40	4	1	1
		20	4	2	1
	3	160	4	2	1
		80	4	2	1
		40	3	1	1
		20	4	2	1
15	1	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1
	2	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	2
	3	160	4	1	1
		80	4	1	1
		40	4	1	1
		20	4	1	1