CONTENTS

LIST OF CONTRIBUTORS vii

PREFACE ix
   Eduardo Salas

1. HOW TO MAKE AUTOMATED SYSTEMS TEAM PLAYERS 1
   Klaus Christoffersen and David D. Woods

2. MULTIMODAL INFORMATION PRESENTATION IN SUPPORT OF HUMAN-AUTOMATION COMMUNICATION AND COORDINATION 13
   Nadine B. Sarter

3. OPERATORS’ AUTOMATION USAGE DECISIONS AND THE SOURCES OF MISUSE AND DISUSE 37
   Hall P. Beck, Mary T. Dzindolet and Linda G. Pierce

   Philip A. Ikomi

5. AUTOMATION AND COGNITION: MAINTAINING COHERENCE IN THE ELECTRONIC COCKPIT 93
   Kathleen L. Mosier

6. IDENTIFYING TRAINING AREAS FOR ADVANCED AUTOMATED AIRCRAFT: APPLICATION OF AN INFORMATION-PROCESSING MODEL 123
   Florian Jentsch, James M. Hitt II and Clint Bowers

7. HUMAN FACTORS ISSUES REGARDING AUTOMATED FLIGHT SYSTEMS 157
   Michael Zuschlag
8. AUTOMATION ISSUES IN REGIONAL AIRLINE OPERATIONS
   Beth Lyall, Greg Harron and Jennifer Wilson 201

9. AUTOMATION AND AGING: ISSUES AND CONSIDERATIONS
   Mustapha Mouloua, Janan Al-Awar Smither,
   Dennis A. Vincenzi and Laura Smith 213
LIST OF CONTRIBUTORS

Hall P. Beck  Appalachian State University, Boone, USA

Clint A. Bowers  University of Central Florida, Orlando, USA

Klaus Christoffersen  The Ohio State University, Columbus, USA

Mary T. Dzindolet  Cameron University, USA

Greg Harron  Research Integrations, Inc., Tempe, USA

James M. Hitt, II  University of Central Florida, Orlando, USA

Philip A. Ikomi  Prairie View A&M University, Prairie View, USA

Florian Jentsch  University of Central Florida, Orlando, USA

Beth Lyall  Research Integrations, Inc., Tempe, USA

Mustapha Mouloua  University of Central Florida, Orlando, USA

Kathleen L. Mosier  San Francisco State University, San Francisco, USA

Linda G. Pierce  Army Research Laboratory, USA
Nadine B. Sarter  The Ohio State University, Columbus, USA
Laura Smith  Embry-Riddle Aeronautical University, USA
Janan Al-Awar Smither  University of Central Florida, Orlando, USA
Dennis A. Vincenzi  Embry-Riddle Aeronautical University, USA
Jennifer Wilson  Research Integrations, Inc., Tempe, USA
David D. Woods  The Ohio State University, Columbus, USA
Michael Zuschlang  National Transportation Systems Center, Cambridge, USA
PREFACE

This is the second volume of the series and focuses entirely on automation. The importances of the topic led me to select this theme. As these chapters document, automation issues are prevalent in many organizations. Industries, agencies, and businesses are relying on automation to solve problems, increase efficiency and productivity, minimize human errors or improve decision-making. All reasonable goals. But automation comes at a cost and in order for it to work, system designers must understand the implication on human performance. And that is what the chapters in this volume explore.

In the first chapter, Christoffersen and Woods argue that two major factors need to be taken into account in order to improve cooperation between humans and automated systems. One, the notion of observability – which suggests that interface designs must be able to effectively provide users with information regarding the current state of the system and the likelihood of the different functions that the system may perform. The other factor, the notion of directability – which suggests that automated systems must be flexible enough to allow users to perform certain corrective actions without having to shutdown the whole system and assume complete control of the situation. Christoffersen and Woods also emphasize that research efforts should focus on improving interface designs as well as on training individuals to better cooperate with such systems.

Similarly, Sarter’s chapter discusses issues related to the effective use of multimodal displays for enhancing human-automation communication and coordination in highly complex event-driven environments. She emphasizes two main areas of investigation that are critical to the effective implementation of multimodal displays (i.e. information presentation and automation coordination), and addresses the main aspects related to each of these. With regard to the former, Sarter points out the importance of focusing on the different affordances and limitations specific to each sensory system, and the context-specific characteristics, which when combined allow for optimal implementation of multimodal displays. With regard to the latter, Sarter describes three main aspects of human-automation coordination (i.e. management of simultaneously presented information, resource allocation, and concurrent feedback loops between human and machine), which when combined allow for optimal implementation of multimodal displays by facilitating time sharing, attention capture, and interruption management.
The third chapter by Beck, Dzindolet, and Pierce, offers a decision-making model to analyze the different parameters associated with the way in which humans choose to use or ignore automation to perform different functions. The authors differentiate between misuse and disuse of automation, and describe three main types of errors associated with each of the two, i.e. recognition, appraisal, and action errors. Beck et al. also point out different variables that contribute to such errors, including machine reliability and consistency, performance feedback, trust, self-confidence, perceived utility, and automation bias.

Similarly, Ikomi presents a study of the effects of automation and mechanical problems on the total number of accidents among commercial airplanes involved in scheduled transportation of passengers and cargo during the period between 1983 and 1999, as reported by the National Transportation Safety Board. Ikomi defines five types of automation-related accidents by differentiating between procedural, mechanical, inadequate-knowledge, mode-utilization, and design problems. Results from this study indicate that automation does not account for a significant proportion of the variance associated with such accidents, whereas mechanical problems do seem to play a significant role. Moreover, Ikomi points out that none of the automation-related accidents lead to losses of lives, and about half of such accidents were caused by mechanical failures of automation equipment.

In Chapter 5, Mosier examines the way in which pilots’ roles in flying have evolved from a correspondence function to a coherence function as a factor of automation. The author points out that this evolution has dramatically changed the cognitive strategies required for pilots to fly aircraft by shifting from highly intuitive responses to more analytical processes.

In Chapter 6, Jentsch, Hitt, and Bowers identify five different aircraft training areas associated with the interaction between flight crews and advanced automated systems based on the information processing (IP) model. These five areas include problems related to mismatches between displays and the capabilities and limitations of human sensory systems; problems related to increased requirements for vigilance, attention-sharing, and attention distribution; problems related to the development of accurate knowledge in long-term memory; problems related to large resource requirements in working memory; and problems related to decision-making, particularly to the selection of actions to achieve certain goals. Results from a cluster analysis indicated that these five areas are closely related to the five different information-processing areas proposed by Wickens. The authors point out that the use of an IP model enables training researchers to effectively partition the different processes associated with the interaction between flight crews and advanced automated systems and allows them to develop solutions, which target specific areas for which they are most helpful.
Next, Zuschlag and Volpe describe the major components of the most widely used automated flight systems, along with a hierarchy of their major functions and modes. They also point out five major problems associated with the interactions between the flight crew and automated systems (i.e. mode awareness, mode prediction, programming errors, programming awkwardness, and diagnosis difficulty) and classify them into three major categories (i.e. operating mode problems, programming problems, and diagnosis difficulty). They further tie these three major categories with Wiener’s three major questions that pilots should address with regard to automated flight systems (i.e. What is it doing now? Why is it doing that? What will it do next?) and propose design solutions to approach relevant problems.

In Chapter 8, Lyall, Harron, and Wilson compare differences between regional and large transportation airlines as they relate to automation and training. They point out that pilots’ turnover rates are much higher among regional airlines and emphasize the need to address training issues to overcome such problems.

The last chapter by Mouloua, Smither, Vincenzi and Smith focuses on automation and aging. They articulate the factors that must be considered when designing systems for older users. They offer a number of suggestions that need to be explored in future studies.

I hope this volume motivates more research in this critical topic in cognitive engineering and probably more importantly, generates a dialogue between those understanding the effects of automation on human performance and systems designers.

Eduardo Salas
University of Central Florida
Series Editor
1. HOW TO MAKE AUTOMATED SYSTEMS TEAM PLAYERS

Klaus Christoffersen and David D. Woods

Interface (noun): an arbitrary line of demarcation set up in order to apportion the blame for malfunctions.


HUMAN-AUTOMATION COOPERATION: WHAT HAVE WE LEARNED?

Advances in technology and new levels of automation have had many effects in operational settings. There have been positive effects from both an economic and a safety point of view. Unfortunately, operational experience, field research, simulation studies, incidents, and occasionally accidents have shown that new and surprising problems have arisen as well. Breakdowns that involve the interaction of operators and computer-based automated systems are a notable and dreadful path to failure in these complex work environments.

Over the years, Human Factors investigators have studied many of the “natural experiments” in human-automation cooperation – observing the consequences in cases where an organization or industry shifted levels and kinds of automation. One notable example has been the many studies of the consequences of new levels and types of automation on the flight deck in commercial transport aircraft (from Wiener & Curry, 1980 to Billings, 1996). These studies have traced how
episodes of technology change have produced many surprising effects on many aspects of the systems in question.

New settings are headed into the same terrain (e.g. free flight in air traffic management, unmanned aerial vehicles, aero-medical evacuation, naval operations, space mission control centers, medication use in hospitals). What can we offer to jump start these cases of organizational and technological change from more than 30 years of investigations on human-automation cooperation (from supervisory control studies in the 1970s to intelligent software agents in the 1990s)?

Ironically, despite the numerous past studies and attempts to synthesize the research, a variety of myths, misperceptions, and debates continue. Furthermore, some stakeholders, aghast at the apparent implications of the research on human-automation problems, contest interpretations of the results and demand even more studies to replicate the sources of the problems.

_Escaping from Attributions of Human Error versus Over-Automation_

Generally, reactions to evidence of problems in human-automation cooperation have taken one of two directions (cf. Norman, 1990). There are those who argue that these failures are due to inherent human limitations and that with just a little more automation we can eliminate the “human error problem” (e.g. “clear misuse of automation . . . contributed to crashes of trouble free aircraft”, La Burthe, 1997). Others argue that our reach has exceeded our grasp – that the problem is over-automation and that the proper response is to revert to lesser degrees of automated control (often this position is attributed to researchers by stakeholders who misunderstand the research results – e.g. (“. . . statements made by . . . Human Factors specialists against automation ‘per se’ ”, La Burthe, 1997). We seem to be locked into a mindset of thinking that technology and people are independent components – either this electronic box failed or that human box failed.

This opposition is a profound misunderstanding of the factors that influence human performance (hence, the commentator’s quip quoted in the epigraph). The primary lesson from careful analysis of incidents and disasters in a large number of industries is that many accidents represent a breakdown in coordination between people and technology (Woods & Sarter, 2000). People cannot be thought about separately from the technological devices that are supposed to assist them. Technological artifacts can enhance human expertise or degrade it, “make us smart” or “make us dumb” (Norman, 1993).

The bottom line of the research is that technology cannot be considered in isolation from the people who use and adapt it (e.g. Hutchins, 1995). Automation
and people have to coordinate as a joint system, a single team (Hutchins, 1995; Billings, 1996). Breakdowns in this team’s coordination is an important path towards disaster. The real lessons of this type of scenario and the potential for constructive progress comes from developing better ways to coordinate the human and machine team – human-centered design (Winograd & Woods, 1997).

The overarching point from the research is that for any non-trivial level of automation to be successful, the key requirement is to design for fluent, coordinated interaction between the human and machine elements of the system. In other words, automation and intelligent systems must be designed to participate in team play (Malin et al., 1991; Malin, 2000).

The Substitution Myth

One of the reasons the introduction of automated technologies into complex work environments can fail or have surprising effects is an implicit belief on the part of designers that automation activities simply can be substituted for human activities without otherwise affecting the operation of the system. This belief is predicated on an assumption that the tasks performed within the system are basically independent. However, when we look closely at these environments, what we actually see is a network of interdependent and mutually adapted activities and artifacts (e.g. Hutchins, 1995). The cognitive demands of the work domain are not met simply by the sum of the efforts of individual agents working in isolation, but are met through the interaction and coordinated efforts of multiple people and machine agents.

Adding or expanding the role of automation changes the nature of the interactions in the system, often affecting the humans’ role in profound ways (one summary is in Woods & Dekker, 2000). For example, the introduction of a partially autonomous machine agent to assist a human operator in a high workload environment is, in many respects, like adding a new team member. This entails new coordination demands for the operator – they must ensure that their own actions and those of the automated agent are synchronized and consistent. Designing to support this type of coordination is a post-condition of more capable, more autonomous automated systems. However meeting this post-condition receives relatively little attention in development projects. The result can be automation which leaves its human partners perplexed, asking Wiener’s (1989) now familiar questions: what is it doing? why is it doing that? what is it going to do next?

As designers, we clearly want to take advantage of the power of computational technologies to automate certain kinds of cognitive work. However, we must realize that the introduction of automation into a complex work
environment is equivalent to the creation of a new cognitive system of distributed human and machine agents and new artifacts. We must also realize the coordination across agents in the system is at least as important as the performance of the individual agents taken in isolation, especially when situations deviate from textbook cases. The attention we give to designing support for this coordination as incidents evolve and escalate can be the determining factor in the success or failure of the human-machine system (Woods & Patterson, 2000).

*How to Design for Coordination: Observability And Directability*

More sophisticated automated systems or suites of automation represent an increase in autonomy and authority (Woods, 1996). Increasing the autonomy and authority of machine agents is not good or bad in itself. The research results indicate that increases in this capability create the demand for greater coordination. The kinds of interfaces and displays sufficient to support human performance for systems with lower levels of autonomy or authority are no longer sufficient to support effective coordination among people and more autonomous machine agents. When automated systems increase autonomy or authority without new tools for coordination, we find *automation surprises* contributing to incidents and accidents (for summaries see Woods, 1993; Woods, Sarter & Billings, 1997; Woods & Sarter, 2000).

The field research results are clear – the issue is not the level of autonomy or authority, but rather the degree of coordination. However, the design implications of this result are less clear. What do research results tell us about how to achieve high levels of coordination between people and machine agents? What is necessary for automated systems to function as cooperative partners rather than as mysterious and obstinate black boxes? The answer, in part, can be stated simply as – Cooperating automation is both observable and directable.

**OBSERVABILITY: OPENING UP THE BLACK BOX**

One of the foundations of any type of cooperative work is a shared representation of the problem situation (e.g. Grosz, 1981; McCarthy et al., 1991). In human-human cooperative work, a common finding is that people continually work to build and maintain a “common ground” of understanding in order to support coordination of their problem solving efforts (e.g. Patterson et al., 1999).

We can break the concept of a shared representation into two basic (although interdependent) parts: (1) a shared representation of the problem state, and (2) representations of the activities of other agents. The first part, shared
representation of the problem situation, means that the agents need to maintain a common understanding of the nature of the problem to be solved. What type of problem is it? Is it a difficult problem or a routine problem? Is it high priority or low priority? What types of solution strategies are appropriate? How is the problem state evolving? The second part, shared representation of other agents’ activities, involves access to information about what other agents are working on, which solution strategies they are pursuing, why they chose a particular strategy, the status of their efforts (e.g. are they having difficulties? why? how long will they be occupied?), and their intentions about what to do next.

Together with a set of stable expectations about the general strategies and behavior of other agents across contexts, mutual knowledge about the current situation supports efficient and effective coordination among problem solving agents (Patterson et al., 1999). Agents can anticipate and track the problem solving efforts of others in light of the problem status and thus coordinate their own actions accordingly. The communicative effort required to correctly interpret others’ actions can be greatly reduced (e.g. short updates can replace lengthy explanations). The ability to understand changes in the state of the monitored process is facilitated (e.g. discerning whether changes are due to a new problem or to the compensatory actions of others). An up to date awareness of the situation also prepares agents to assist one another if they require help.

Notice how much of the knowledge discussed here is available at relatively low cost in “open” work environments involving multiple human agents. For example, in older, hardwired control centers, individual controllers can often infer what other controllers are working on just by observing which displays or control panels they are attending to. In the operating room, surgical team members can observe the activities of other team members and have relatively direct, common access to information about the problem (patient) state. The open nature of these environments allows agents to make intelligent judgments about what actions are necessary and when they should be taken, often without any explicit communication. However, when we consider automated team members, this information no longer comes for free – we have to actively design representations to generate the shared understandings which are needed to support cooperative work.

Data Availability Does Not Equal Informativeness

Creating observable machine agents requires more than just making data about their activities available (e.g. O’Regan, 1992). As machine agents increase in complexity and autonomy, simple presentations of low-level data become insufficient to support effective interaction with human operators. For example,
many early expert systems “explained” their behavior by providing lists of the individual rules which had fired while working through a problem. While the data necessary to interpret the system’s behavior was, in a literal sense, available to operators, the amount of cognitive work required to extract a useful, integrated assessment from such a representation was often prohibitive. A more useful strategy was to provide access to the intermediate computations and partial conclusions that the machine agent generated as it worked on a problem. These were valuable because they summarized the machine agent’s conception of the problem and the bases for its decisions at various points during the solution process.

In general, increases in the complexity and autonomy of machine agents requires a proportionate increase in the feedback they provide to their human partners about their activities. Representations to support this feedback process must emphasize an integrated, dynamic picture of the current situation, agent activities, and how these may evolve in the future. Otherwise, mis-assessments and miscommunications may persist between the human and machine agents until they become apparent through resulting abnormal behavior in the process being controlled. For example, the relatively crude mode indicators in the current generation of airliner cockpits have been implicated in at least one major air disaster. It is clearly unacceptable if the first feedback pilots receive about a miscommunication with automation is the activation of the ground proximity alarm (or worse).

Human agents need to be able to maintain an understanding of the problem from the machine agent’s perspective. For instance, it can be very valuable to provide a representation of how hard the machine agent is having to work to solve a problem. Is a problem proving especially difficult? Why? If the automated agent has a fixed repertoire of solution tactics, which have been tried? Why did they fail? What other options are being considered? How close is the automation to the limits of its competence? Having this sort of information at hand can be extremely important to allow a human agent to intervene appropriately in an escalating critical situation.

Providing effective feedback to operators in complex, highly automated environments represents a significant challenge to which there are no ready-made solutions. Answering this challenge for the current and future generations of automation will require fundamentally new approaches to designing representations of automation activity (e.g. Sarter, 1999; Sklar & Sarter, 1999; Nikolic & Sarter, 2001). While the development of these approaches remains to be completed, we can at least sketch some of the characteristics of these representation strategies (Woods & Sarter, 2000). The new concepts will need to be:
• Event-based: representations will need to highlight changes and events in ways that the current generation of state-oriented display techniques do not.

• Future-oriented: in addition to historical information, new techniques will need to include explicit support for anticipatory reasoning, revealing information about what should/will happen next and when.

• Pattern-based: operators must be able to quickly scan displays and pick up possible abnormalities or unexpected conditions at a glance rather than having to read and mentally integrate many individual pieces of data

DIRECTABILITY: WHO OWNS THE PROBLEM?

Giving human agents the ability to observe the automation’s reasoning processes is only one side of the coin in shaping machine agents into team players. Without also giving the users the ability to substantively influence the machine agent’s activities, their position is not significantly improved. One of the key issues which quickly emerges in trying to design a cooperative human-machine system is the question of control. Who is really in charge of how problems are solved? As Billings (1996) pointed out, as long as some humans remain responsible for the outcomes, they must also be granted effective authority and therefore ultimate control over how problems are solved. Giving humans control over how problems are solved entails that we, as designers, view the automation as a resource which exists to assist human agents in the process of their problem solving efforts.

While automation and human activities may integrate smoothly during routine situations, unanticipated problems are a fact of life in complex work environments such as those where we typically find advanced automation. It is impossible in practice, if not in principle, to design automated systems which account for every situation they might encounter. While entirely novel problems may be quite rare, a more common and potentially more troublesome class of situations are those which present complicating factors on top of typical, “textbook” cases (cf. studies of brittleness of automated systems include Roth et al., 1987; Guerlain et al., 1996; Smith et al., 1997). These cases challenge the assumptions on which the pre-defined responses are based, calling for strategic and tactical choices which are, by definition, outside the scope of the automation’s repertoire. The relevant question is, when these sorts of problems or surprises arise, can the joint system adapt successfully?
Traditionally, one response to this need has been to allow human operators to interrupt the automation and take over a problem manually. Conceiving of control in this way, an all-or-nothing fashion, means that the system is limited to operating in essentially one of two modes – fully manual or fully automatic. This forces people to buy control of the problem at the price of the considerable computational power and many potentially useful functions which the automation affords. What is required are intermediate, cooperative modes of interaction which allow human operators to focus the power of the automation on particular sub-problems, or to specify solution methods that account for unique aspects of the situation which the automated agent may be unaware of. In simple terms, automated agents need to be flexible and they need to be good at taking direction.

Part of the reason that directability is so important is that the penalties for its absence tend to accrue during those critical, rapidly deteriorating situations where the consequences can be most severe. One of the patterns that we see in the dynamic behavior of complex human-machine systems during abnormal situations is an *escalation* in the cognitive and coordinative demands placed on human operators (Woods & Patterson, 2000). When a suspicious or anomalous state develops, monitoring and attentional demands increase; diagnostic activities may need to be initiated; actions to protect the integrity of the process may have to be undertaken and monitored for success; coordination demands increase as additional personnel/experts are called upon to assist with the problem; others may need to be informed about impacts to processes under their control; plans must be modified, contingencies considered; critical decisions need to be formulated and executed in synchronization with other activities. All of this can occur under time pressure (Klein et al., 2000).

These results do not imply that automation work only as a passive adjunct to the human agent. This is to fall right back into the false dichotomy of people versus automation. Clearly, it would be a waste of both humans’ and automation’s potential to put the human in the role of micro-managing the machine agent. At the same time however, we need to preserve the ability of human agents to act in a strategic role, managing the activities of automation in ways that support the overall effectiveness of the joint system. As was found for the case of observability, one of the main challenges is to determine what levels and modes of interaction will be meaningful and useful to practitioners. In some cases human agents may want to take very detailed control of some portion of a problem, specifying exactly what decisions are made and in what sequence, while in others they may want only to make very general, high level corrections to the course of the solution in progress. Accommodating all of these possibilities is difficult and requires careful iterative analysis of the interactions.
between system goals, situational factors, and the nature of the machine agent. However, this process is crucial if the joint system is to perform effectively in the broadest possible range of scenarios (Roth et al., 1997; Dekker & Woods, 1999; Guerlain et al., 1999; Smith et al., 2000; Smith, in press).

In contrast to this, technology-driven designs tend to isolate the activities of humans and automation in the attempt to create neatly encapsulated, pseudo-independent machine agents. This philosophy assumes that the locus of expertise in the joint human-machine system lies with the machine agent, and that the human’s role is (or ought to be) largely peripheral. Such designs give de facto control over how problems are solved to the machine agent. However, experience has shown that when human agents are ultimately responsible for the performance of the system, they will actively devise means to influence it. For example, pilots in highly automated commercial aircraft have been known to simply switch off some automated systems in critical situations because they have either lost track of what the automation is doing, or cannot reconcile the automation’s activities with their own perception of the problem situation. Rather than trying to sort out the state of the automation, they revert to manual or direct control as a way to reclaim understanding of and control over the situation. The uncooperative nature of the automated systems forces the pilots to buy this awareness and control at the price of abandoning the potentially useful functions that the automation performs, thus leaving them to face the situation unaided.

Whither Automated Agents? Invest in Design for Team Play

Repeatedly, performance demands and resource pressures lead mission organizations to invest in increasing the autonomy and authority of automated systems. Because of unquestioned assumptions that people and automated systems are independent and inter-changeable, organizations fail to make parallel investments in design for observability and directability. Often in the process of recruiting resources for new levels of automation, advocates vigorously promote the claim that the more autonomous the machine, the less the required investment in team play and the greater the savings for the organization.

The operational effects of this pattern of thinking are strikingly consistent. Inevitably, situations arise requiring team play; inevitably, the automation is brittle at the boundaries of its capabilities; inevitably, coordination breakdowns occur when designs fail to support collaborative interplay; and inevitably, operational personnel must scramble to work around clumsy automation which is ill-adapted to the full range of problems or to working smoothly with
other agents. Meanwhile, cycling in the background, commentators from various perspectives bicker about crediting one or another agent as the sole cause of system failures (Woods & Sarter, 2000).

We have no need to witness or document more of these natural experiments in strong, silent, difficult to direct automation. Experience has provided us with ample evidence for the shallowness, error, and sterility of these conventional beliefs. If we simply drop the blinders of the Substitution Myth, the scene comes into clear focus (Woods & Tinapple, 1999). The analysis of past natural experiments reveals ways to go forward. Because of increasing capabilities of automated systems, the design issue is collaboration within the joint human-machine system as this joint system copes with the variety and dynamics of situations that can occur. For this joint human-machine system to operate successfully, automated agents need to be conceived and designed as “team players”. Two of the key elements needed to support this coordinated cognitive work are observability and directability.

**SUMMARY**

When designing a joint system for a complex, dynamic, open environment, where the consequences of poor performance by the joint system are potentially grave, the need to shape the machine agents into team players is critical. Traditionally, the assumption has been that if a joint system fails to perform adequately, the cause can be traced to so-called “human error.” However, if one digs a little deeper, they find that the only reason many of these joint systems perform adequately at all is because of the resourcefulness and adaptability that the human agents display in the face of uncommunicative and uncooperative machine agents. The ability of a joint system to perform effectively in the face of difficult problems depends intimately on the ability of the human and machine agents to coordinate and capitalize upon the unique abilities and information to which each agent has access.

For automated agents to become team players, there are two fundamental characteristics which need to be designed in from the beginning: observability and directability. In other words, users need to be able to see what the automated agents are doing and what they will do next relative to the state of the process, and users need to be able to re-direct machine activities fluently in instances where they recognize a need to intervene. These two basic capabilities are the keys to fostering a cooperative relationship between the human and machine agents in any joint system.
NOTE

1. Recall that intelligent automation has often been introduced as an attempt to replace “inefficient” or “error-prone” human problem solvers.

REFERENCES


INTRODUCTION

Designers of human-machine interfaces have focused for a long time on the development and refinement of visual and, to a more limited extent, auditory displays. Only in recent years has a significant and rapidly increasing interest in multimodal systems emerged. To a large extent, this interest was triggered by the evolution of virtual reality (VR) technologies. Multimodal interfaces have been developed that support movement through, and interaction with, these environments and thus increase the perceived immersion in virtual worlds. Multimodal technologies have been proposed for other purposes and domains as well. They involve the potential for making computing accessible to a wider range of users, including people from different age groups, with different cognitive styles, and those with sensory or motor impairments. Multimodal interfaces play an important role in the development of VR-based training.
systems (such as surgical training systems for minimally invasive surgery where tactile and force feedback is needed), and they are being developed in support of mobile computing and telerobotics.

In this chapter, we will discuss yet another reason for developing multimodal interfaces: the introduction of increasingly complex and autonomous systems to a variety of work domains where operators need to collaborate with other human and machine agents. The introduction of these systems has been a mixed blessing. While benefits such as an increased precision and efficiency of operations were achieved, they appear to accrue primarily in situations where automated systems perform tasks on their own, i.e. without operator involvement. In other situations, where humans and machines need to cooperate and coordinate their activities, unexpected problems are sometimes encountered. To a large extent, these difficulties can be explained by the fact that modern systems have evolved from passive tools to autonomous agents without acquiring the communicative skills required for negotiating and coordinating goals and actions with their human colleagues in a timely and efficient manner (Norman, 1990). Data on the status and behavior of these systems is generally available but it is not presented in a form that matches human information needs and processing abilities and limitations. This can create problems, especially in the case of automation behavior that occurs independent of, and sometimes in conflict with, operator instructions. The result is often ‘automation surprises’ and mode errors where operator behavior is appropriate for the presumed but not for the actual status of the system (Sarter, Woods & Billings, in press; Woods et al., 1994).

In aviation, for example, pilots’ difficulties with maintaining awareness of the current and future status and behavior of their automated flight deck systems have been the focus of many research efforts (e.g. Abbott et al., 1996; Javaux & DeKeyser, 1998; Sarter & Woods, 1994, 1995, 1997, 2000; Wiener, 1989). These studies have shown that pilots tend to miss transitions between automation states and behaviors when these transitions occur in the absence of immediately preceding pilot input and are therefore difficult to predict and monitor.

A variety of top-down and bottom-up influences contribute to the observed difficulties with monitoring and maintaining mode awareness. Gaps and misconceptions in pilots’ mental models play a role as they can lead to inappropriate expectations of system behavior and thus inappropriate attention allocation, both in terms of timing and focus (e.g. Wiener, 1989; Sarter & Woods, 1994, 1997; Mumaw et al., 2001a, b). Inattentional blindness (e.g. Rensink, O’Regan & Clark, 1997) is another possible reason for missing unexpected mode transitions. A person who is engaged in an attentionally demanding task or whose attention is focused on a particular (display) location, is likely to miss other, unexpected objects or events (such as uncommanded mode transitions). Gaps in system
knowledge and understanding as well as inattentional blindness exemplify top-
down influences on attention allocation.

Bottom-up processes also affect monitoring performance. Change blindness
(Simons, 2000) is one example. It refers to the fact that people are surprisingly
poor at noticing the appearance of, or even large changes to, objects when
these onsets or changes occur simultaneously with a transient such as an eye
movement or global or local disruptions in a display. On modern flight decks,
change blindness may be experienced when mode indications, which are
embedded in a highly dynamic data-rich display (the so-called Primary Flight
Display (PFD)), change concurrently with other elements on the PFD (e.g.

Ultimately, the effectiveness and timeliness of attention allocation and thus
monitoring is determined by the interaction of top-down and bottom-up
processes in the overall perceptual cycle (Neisser, 1976). For example, a pilot
sometimes cannot avoid focusing his/her attention on some task other than
monitoring the automation. This leads to an increased risk of missing unexpected
mode transitions. However, this risk can be reduced by enabling the automation
to provide system-generated attentional guidance. The system can highlight a
change or event and thus capture and guide the pilot’s attention to relevant
indications. External attentional guidance is essential for human-machine coor-
dination in any event-driven domain, and it will become even more important
as systems continue to become more complex and autonomous.

To date, most automated systems do not possess the communication and
coordination skills required to support this function effectively. One reason for
this shortcoming is the almost exclusive reliance in interface design on visual
information presentation. This trend prevents the exploitation of affordances of
other sensory channels that are available to human operators, such as touch,
smell, and taste. Distributing cues across these modalities can support parallel
access to, and processing of, information – an ability that is critical in data-rich
environments such as the flight deck. Our various senses are tuned to different
kinds of cues and thus combining modalities can result in mutual compensation
for the limitations associated with each channel (Oviatt, 2000). Also, and very
importantly, multimodal information presentation can support coordinative
functions, such as attention and interruption management (Sarter, 2000).

In this chapter, some of the potential uses, benefits, and limitations of multi-
modal information presentation will be discussed. Our focus will be on passive
multimodal interfaces, i.e. interfaces that present information to a user rather
than allow for user input to a system. This merely reflects our own primary
research interests, i.e. enabling modern technologies to play a more active
role in human-machine communication and coordination. It is by no means a
judgment of the relative importance and benefits of work on active versus passive displays. In fact, the creation of robust multimodal systems will ultimately require researchers in this field to consider the overall cycle of system input and output rather than focus on either component in isolation.

MULTIMODAL INFORMATION PRESENTATION: AFFORDANCES, LIMITATIONS, AND THE CONTEXT-SENSITIVE INTEGRATION AND ADAPTATION OF MODALITIES

Multimodal information presentation is not simply a means of enhancing the bandwidth of our information processing capacity. It also represents a promising approach to improving the quality of information presentation and supporting coordinative functions such as attention and interruption management. Several authors have argued that, to achieve these goals, “different display and control requirements cannot be arbitrarily assigned to auditory and speech [and other; the author] channels” (e.g. Wickens, Sandry & Vidulich, 1983). Instead, they propose that theory-based principles need to be developed that inform when, and for what purposes, different modalities should be employed, and how they can be combined effectively. This seems to suggest a static approach to modality assignments, which is problematic for a number of reasons. Before we discuss those reasons, however, a brief overview of some of the affordances and limitations of vision, hearing, touch, and smell will be provided.

As mentioned earlier, visual information presentation is heavily emphasized in today’s interface designs. For the most part, these interfaces rely on foveal vision, which seems appropriate for the presentation of complex graphics and, in general, for conveying large amounts of detailed information. A related advantage of visual information is its potential for permanent presentation, which affords delayed and prolonged attending. Also, vision does not require us to be as close to a (potentially dangerous) source of information as, for example, taste or smell.

Current interfaces make deliberate use of peripheral vision to a much lesser extent. While the two channels – foveal and peripheral vision – do not, in the strict sense, represent separate modalities, they are associated with different options and constraints. Peripheral vision is well suited for detecting motion, luminance changes, and the appearance of new objects. In contrast, foveal vision supports the recognition of objects or details. Peripheral vision is assumed to determine the destination of our next saccade (McConkie, 1983). In that sense, it represents an early orientation mechanism that can be utilized by designers to help operators attend to a relevant location or critical information at the right
time. A potential problem with peripheral visual feedback is that the visual field changes dynamically in response to contextual factors. With increasing foveal taskloading, for example, visual attention begins to focus on information in the center of a display at the expense of information presented in peripheral vision – a phenomenon called “attentional narrowing”. Recent research suggests, however, that practice in divided attention tasks can reduce the vulnerability to attentional narrowing (Williams, 1995).

The auditory channel differs from vision along several dimensions. First, it is omnidirectional, thus allowing for information to be picked up from any direction. Secondly, auditory information presentation is transient. This potential limitation is compensated for by a longer short-term storage of auditory (as opposed to visual) information so that it can be processed with some delay. Finally, since it is impossible for us to “close our ears,” auditory displays tend to be intrusive and are therefore often reserved for alerting functions.

A currently underutilized channel for presenting information is the haptic sense. Haptic sensory information can take the form of either tactile or kinesthetic cues. Tactile feedback is presented to the skin in the form of force, texture, vibration, and thermal sensations. Kinesthetic, or proprioceptive, cues convey information about body position and motion and the forces acting on the body. The sense of touch shares a number of properties with the auditory channel. Most importantly, cues presented via these two modalities are transient in nature. Also, like vision and hearing, touch allows for the concurrent presentation and extraction of several dimensions, such as frequency and amplitude in the case of vibrotactile cues. Touch differs from vision and hearing in that it is capable of both sensing and acting on the environment.

Another possible form of information presentation – olfactory interfaces – is the least developed within the field of multimodal interaction for a number of reasons. First, useful applications appear to be limited although this is changing with the advent of advanced VR technologies. Difficulties with implementing delivery systems for olfactory cues have slowed down progress in this area as well. One of the main challenges is the need for odor storage and for controlling the breathing space for the individual being presented with these cues, which is especially difficult in real-world environments. Advances in the development of olfactory interfaces are desirable because odors can support numerous functions such as conveying high-level assessments of a situation (including alerting to life-threatening situations), increasing vigilance, decreasing stress, and improving retention and recall of learned material. The olfactory sense plays an important role in arriving at decisions quickly in an experiential, rather than reflective, manner. Even more so than other modalities, olfaction involves large inter- and intrapersonal differences (Vroon, 1997).
It is important to note that the effectiveness of each of the above sensory channels for information presentation depends heavily on context. It is affected by the mental state of a person and by interactions and interference between modalities. For example, the speed of detection and response to different modalities varies with a person’s state. It appears to shift from the visual to the auditory channel if subjects are in state of aversive arousal (Johnson & Shapiro, 1989). Another important crossmodal phenomenon is visual dominance (Wickens, 1992), which counteracts the natural tendency of humans to orient their attention to stimuli in the auditory and tactile modality. Visual dominance is observed when visual stimuli appear at about the same frequency and/or are of the same relevance to the person as the other cues. Other factors determining the effectiveness of presenting information via various sensory channels – such as crossmodal interference, modality expectations, and the modality shifting effect – will be discussed later in the section on time-sharing.

The fact that context determines to a considerable extent the effectiveness of information presentation via different channels calls into question any static assignments of modalities to types and sources of information. Instead, similar to the allocation of tasks and functions to human and machine agents, the use of different modalities will need to be adaptive; it needs to consider the abilities and preferences of individual operators, environmental conditions, task requirements/combinations, and degraded operations, which may render the use of certain channels ineffectual. In developing an adaptive approach to multimodal information presentation, it will be critical to avoid that the flexibility afforded by that approach is achieved at the cost of imposing new interface management and coordination tasks that could divert the operator’s attention away from ongoing critical tasks and events. Dynamic adaptation of modalities requires that human and machine agents are supported in rapidly changing and combining information presentation modes and in maintaining “modality awareness”.

One possible challenge for adaptive modality changes is the so-called modality-shifting effect (MSE). It refers to the fact that, if people have just responded to a cue in one modality, they tend to be slower to respond to a subsequent cue in a different modality (Spence & Driver, 1996). This could create difficulties in operational environments that involve high risk and require rapid operator responses (such as the military domain). Note, however, that in most domains, the delays that were observed in laboratory studies seem too short to be of operational significance.

A dynamic approach to modality assignments in collaborative environments also requires the development of effective protocols for coordinating modality
changes. When and why are changes required, feasible, and acceptable? Who should be allowed to initiate, and make a final decision about, such changes? How can modality changes be negotiated and coordinated effectively?

In addition to supporting adaptive modality changes that are initiated by other human agents, it will be important to explore options for auto-adaptive changes in information presentation. These will be critical especially under high workload high-tempo conditions when operators need to focus on the task at hand and cannot afford to consider a modality change nor handle its negotiation and implementation. One possibility is for some automated system – a kind of “modality coordinator” – to monitor environmental conditions such as noise or visibility and change automatically and accordingly the modality of information presentation. Such a system may also be able to monitor currently utilized communication channels and synchronize the presentation of disparate data sources by determining available and appropriate modalities. Some of the questions that need to be addressed in this area include: (a) how can we support the operator in noticing and adapting to an automation-induced modality change? (b) how can he/she reverse or modify the change if needed? and (c) how can human and system coordinate changes in information presentation?

Effective adaptation to modality changes is one important prerequisite for a robust multimodal system. The ability to process information that is presented concurrently via different modalities is another. Concurrent multimodal information presentation can serve: (a) to provide information on various concurrent processes/events or (b) to present information on various aspects of the same event or process. Presenting information on various concurrent processes requires timesharing and allows for parallel tracking of multiple sources of information. In that sense, it can support building common ground and coordinating or synchronizing the activities of multiple agents.

In contrast, using various modalities to present information on different aspects of the same event can support effective joining of this information to form a coherent picture of a situation. This use of multimodal information presentation seems to be in line with users’ natural organization of multimodal interaction, which tends to be oriented towards complementarity rather than redundancy (Oviatt, 1999). It enables the creation of pattern-based representations, which are important for diagnosing and detecting changes and events.

As mentioned earlier, multimodal interfaces are not simply a means of increasing the bandwidth of our processing capacity and improving the quality of information presentation. They not only contribute to a more natural form of human-computer interaction but can serve specifically to support coordinative functions in human-human and human-machine teams. The following section
will discuss some important coordination mechanisms and present empirical evidence that shows how coordination can benefit from multimodal information presentation.

MULTIMODAL SYSTEMS IN SUPPORT OF HUMAN-AUTOMATION COORDINATION

Coordination is the process of synchronizing the interdependent activities of various agents. Three important components of coordination are: (a) the management of simultaneity constraints, (b) effective resource allocation, and (c) the communication of intermediate results and difficulties with task performance (Malone, 1990).

The management of simultaneity constraints and effective resource allocation become relevant, for example, when several tasks either need to be performed concurrently or cannot occur at the same time. In the former case, conflicts between tasks and therefore performance decrements are likely unless resource competition can be avoided. One way to achieve this goal was suggested by the original version of multiple resource theory (Wickens, 1984), which proposes that modalities are associated with separate attentional resources. By distributing tasks across sensory channels, more effective timesharing should be possible.

Another important aspect of coordination is the communication of both intermediate results and difficulties with task performance. This relates to the need for human and machine agents to engage in interactive give and take. In particular, highly autonomous systems need to go beyond responding to user requests or inquiries. In the interest of effective coordination, they need to ask clarification questions on their own, report on progress, and submit intermediate results to their human counterparts. The challenge associated with this requirement is that most modern technologies are not capable of recognizing the costs of interaction. In other words, they cannot determine whether the human is engaged in some important task that should not be interrupted. At best, they can support the operator in making that decision on his/her own by providing information on the nature of an interruption. Multimodal information presentation can be useful in this context as it can serve to: (a) capture attention and alert the operator to the fact that something may require his/her attention, and (b) present interruption-related information in a way that does not conflict with ongoing tasks and thus support effective interruption management.

The following sections will discuss in more detail how multimodal systems can support these three coordination functions: (a) timesharing, (b) attention
capture, and (c) interruption management. Empirical evidence will be presented for each case.

(A) Multimodal Information Presentation in Support of Time-Sharing

One hope associated with the development of multimodal information presentation is that it will support operators in timesharing, i.e. in the parallel processing of information and the simultaneous performance of multiple tasks. While there is considerable empirical evidence for such performance benefits, it is not clear whether these benefits are indeed related to central processing mechanisms, or whether they are achieved due to peripheral factors, such as the avoidance of visual scanning costs (Wickens & Liu, 1988).

Recent research suggests that people do not possess entirely independent resources for processing information in separate modalities. Spence and Driver (1996) have provided an excellent overview of these findings. For example, expecting a cue to appear in a certain modality increases the detection rate and reduces the response time to that stimulus. Post and Chapman (1991) have shown a 7–10% decrease in reaction time to visual and tactile stimuli if they appeared in the expected modality. At the same time, however, such expectations will hurt detection performance for cues that appear in a modality other than the expected. Again, Post and Chapman (1991) report an increase of 16–22% in the reaction time to cues in the unexpected modality. Overall, they propose a gradient model for modality expectations where detection performance is best for cues in an expected modality, worst for those in an unexpected modality, and intermediate for cues if no specific modality expectation exists.

Modality expectations may be created by the observed or known higher frequency of cues in one modality compared to others. Increased attention to one modality may also be the result of the actual or perceived importance of cues in that modality. Crossmodal spatial links can create a related problem when, for example, the expectation of an auditory cue in a certain location leads to a shift of visual attention towards that same location (Spence & Driver, 1996).

(B) Multimodal Information Presentation in Support of Attention Capture

One important prerequisite for playing an active and effective role in coordination is the ability to attract attention in a timely and reliable fashion. Observed breakdowns in pilot-automation interaction are, to some extent, symptoms of a failure to enable highly automated flight deck systems to perform this function. As mentioned earlier, these systems rely heavily and increasingly on focal visual
attention, and they provide feedback primarily on system status rather than changes and events.

One possible way to address this problem is the introduction of auditory cues, which has been proposed under the following circumstances: when a message is short and simple, will not be referred to later, calls for immediate action, or the visual system is overburdened (Deathridge, 1972). The latter problem is often encountered on modern flight decks and may be overcome by introducing auditory cues, which are omni-directional, thus not requiring the pilot to monitor a particular display or to maintain a particular head orientation (Wickens, 1992). Also, as discussed earlier, input to the auditory channel is transient but believed to be available for review for a slightly longer period of time than visual information. During that time, a deliberate switch of attention can be made to the information (Wickens, 1992).

Despite these advantages, there is considerable opposition to presenting auditory automation feedback on modern flight decks. Most notably, pilots complain about the already heavy use of auditory cues in the cockpit, which, if further increased, may result in “auditory clutter”. Also, as long as voice communication is used for air-ground communication, auditory system messages can interfere with the perception of air traffic control instructions. Finally, auditory cues tend to be reserved for highly critical messages related to alarms and warnings. Mode transitions do not fall into this category. They may have been anticipated and/or noticed already by the pilot and therefore do not justify the high level of alerting associated with most auditory cues. We will return to this subject in our discussion of interruption management.

Another, more promising, approach to supporting mode awareness is to present automation-related information in peripheral vision (e.g. Venturino & Rinalducci, 1986), which has been hypothesized to be a distinct information processing channel (Leibowitz, 1988). Peripheral vision has several properties that make it a promising candidate for attention capture and guidance. Most importantly, the periphery is very effective at detecting motion and luminance changes. This adaptive feature of peripheral vision makes it useful for detecting potentially interesting objects that may warrant subsequent focal attention. Information that is presented in the periphery is obtained with little or no conscious effort (Jonides, 1981), making peripheral vision a resource-economical channel. Empirical evidence from physiological, functional, and cognitive research seems to support the dual nature of vision which allows focal and peripheral inputs to be processed in parallel (Christensen et al., 1986).

Like auditory cues, peripheral visual feedback involves a number of potential drawbacks. Most notably, breakdowns in monitoring could occur due to the earlier mentioned narrowing of the functional field of view under conditions of
increasing stress and cognitive task loading (Leibowitz & Appelle, 1969). Also, in domains like aviation where operators are required to divide their attention widely across a large display area, it is difficult to ensure that indications will reliably appear in their peripheral visual field.

Yet another possible approach to the attention management challenge in multiple-dynamic domains is the introduction of currently underutilized tactile cues. This approach has been promoted by several researchers in the aviation domain (e.g. Endsley, 1988; Furness, 1986; Hawkes, 1962). Limited research exists on how to make the most effective use of this form of information presentation. Potential limitations of tactile cues, such as masking effects (e.g. Chapman, Bushnell, Miron, Duncan & Lund, 1987; Kirman, 1986; Post & Chapman, 1991; Verrillo, 1983), have yet to be studied in more detail. Also, most tactile research to date has been conducted in laboratory settings (e.g. Butter, Buchtel & Sanctucci, 1989; Shiffrin, Craig & Cohen, 1973), and it is not clear that those findings scale up to more complex data-rich domains. To date, few research efforts have examined tactile feedback within an aviation context (e.g. Ballard & Hessinger, 1954; Gilliland & Schlegel, 1994; Zlotnik, 1988). These studies have used touch primarily for providing navigational guidance and for alerting purposes.

Two experiments were recently conducted in our laboratory to assess the effectiveness, as well as potential costs and limitations, of using peripheral visual and tactile cues for attention capture in case of unexpected discrete events. Subjects in these studies were instructor pilots who were asked to fly a scenario on a desktop simulator of a modern glass cockpit aircraft. In addition to their flight-related tasks, which varied in terms of number and difficulty, pilots had to detect unexpected changes in the status of an automated system (so-called mode transitions) as well as other events (such as traffic) that occurred either in isolation or concurrently with a mode transition.

The first study compared the effectiveness of current visual automation feedback (so-called flight mode annunciations – FMAs) with two types of peripheral visual cues for indicating the occurrence of uncommanded and unexpected transitions (Nikolic & Sarter, 2001). Mode transitions are currently indicated by the appearance of a thin outline box around the corresponding alphanumeric mode indication. The new active mode is displayed, and the box disappears after 10 seconds. In our study, the first improved peripheral display was a more salient version (a solid colored box) of the current transition indication in the same location. The second peripheral display was an ambient strip, which signaled mode transitions by the onset of a thin colored band that was located at the bottom of the monitors in front of the pilot. It subtended 60° of visual angle to increase its chances of being noticed by the pilot.
The second study (Sklar & Sarter, 1999) examined the effectiveness of using a dynamic tactile display (i.e. a display that presents patterns of stimulation to passive skin surfaces; see Craig & Sherrick, 1982) for the same purpose. Vibrotactile cues were presented via tactors to the subject’s wrist whenever a mode transition occurred. Participants were presented either with tactile cues only or with tactile cues in combination with the current visual feedback (FMA).

Both peripheral visual and, even more so, tactile cues resulted in faster response times and significantly improved detection performance for the experimenter-induced mode transitions (see Fig. 1) without affecting the performance of concurrent tasks such as flight path tracking or the detection of other important events. This effect was most pronounced when pilots had to handle competing attentional demands related to the management of their automated systems (see results for Auto mgmt phase in Fig. 1).

Pilots receiving the improved peripheral visual cues performed better than the baseline group but still missed a considerable number of events. A recent follow-up study suggests that this result may be explained, in part, by the fact that the proposed attention capture power of peripheral visual cues (e.g. Jonides, 1981, Yantis & Jonides, 1990) is affected considerably by the background or context in which these cues appear (Nikolic, Orr & Sarter, 2001). Participants in the study had to perform an externally-paced visually demanding task on one monitor while, at the same time, detecting targets that appeared on an adjacent monitor. These targets (green boxes similar to those used in the above study) appeared either against a solid black background (control) or in the context of other display objects (round-dial gauges and number displays). The surrounding objects were either static (stat) or dynamic (dyn), and they were either white (mono) or of the same color (color) as the target boxes. If onsets were embedded in a dynamic background involving objects of similar color as the onset itself (which is the case with mode annunciations on current flight decks), a significant decrease in detection performance was observed (see Fig. 2). This effect was particularly strong for targets that were presented at a larger distance from the primary visual task.

The findings from the above studies illustrate the potential of multimodal information presentation for addressing observed difficulties with attention capture, and thus human-automation communication and coordination, in data-rich highly dynamic environments. It is important to keep in mind, however, that effective coordination requires more than reliable attention capture. Not every event warrants an immediate re-orientation of the operator’s attention. Instead, the decision about when to respond to an interruption depends on the actual and perceived demands on the operator and the relative importance of the interruption. The problem is that even highly advanced automated
systems do not necessarily have access to the contextual information that is required to make this determination. At best, they can support operators in making interruption management decisions themselves.

\( (C) \) Multimodal Information Presentation in Support of Interruption Management

Interruptions and the preoccupation with one task at the expense of another negatively affect performance (e.g. Detweiler, Hess & Phelps, 1994; Gillie & Broadbent, 1989) and often play a role in incidents and accidents (e.g. Dismukes, Young & Sumwalt, 1998; Latorella, 1999; Dornheim, 2000). One of the most

Fig. 1. Detection Performance for Unexpected Mode Transitions for a Baseline Group (FMA – receiving currently available feedback) and two groups each receiving either peripheral visual (enhanced FMA and ambient strip) or vibrotactile (tactile+FMA and tactile only) feedback. Subjects had to perform either no concurrent task (none), an easy versus difficult tracking task, or automation management tasks (auto mgmt).
Fig. 2. Detection Performance for Unexpected Targets as a Function of Display Context and Target Displacement.
frequently observed problems following an interruption are lapses, i.e. failures to perform an intended but deferred action. One possible way to address this problem is to provide effective reminder mechanisms. An example of this approach in the aviation domain are so-called electronic checklists on modern flight decks. These checklists keep track of which items have already been completed and remind pilots of checks that still need to be done.

Another approach to interruption management could be to have an automated system filter and schedule interruptions for the operator. Once a potential interruption is detected, the system could determine the current workload and attentional demands on the operator and, if necessary, delay or suppress delivery of the interruption message or task. This technology-centered approach to interruption management appears not only inadvisable but also infeasible given the above-mentioned level of contextual blindness of many modern systems.

Voice loops represent another, more effective, means of supporting interruption management. They allow all involved parties to gauge each other’s interruptability by listening in on conversations and thus tracking the demands on each participant (Patterson, Watts-Perotti & Woods, 1999). A similarly pro-active and human-centered (Billings, 1996) approach to interruption management is to assist operators in determining on their own whether or not, and how fast, they should respond to an interruption. This approach can help reduce the number of unnecessary interruptions of ongoing tasks and lines of reasoning rather than address the potentially negative consequences of interruptions after-the-fact.

One of the early examples of this approach are so-called likelihood alarm displays (Sorkin, Kantowitz & Kantowitz, 1988) where the likelihood of an event is computed by an automated monitoring system and then encoded in an alerting signal. For example, an operator may be informed that a certain system failure is possible, probable, likely, or certain. Under high workload conditions, this information is more effective than traditional binary alarm signals in helping operators decide whether and when to interrupt ongoing tasks to attend to the problem without imposing undue attentional demands.

Interruptions are not always caused by an alarm. They can also occur as the result of externally imposed tasks or when another person or an automated system is initiating an interaction to inform the operator of progress on, or difficulties with, performing a task. Other agents may need to negotiate changing goals and intentions or request information or approval from the human operator (Ball et al., 1997). These types of interruptions will become increasingly likely as modern technologies become more autonomous.

Operators can be supported in handling these interruptions more effectively by providing them with information about the nature of the interruption. Various
types of information may be relevant, depending on task and task context. These can include the source and urgency of the interruption, the time remaining and the time required to handle the interruption, and the type of information processing required by the interruption (for example, spatial versus verbal information). The decision whether to attend to an interruption also likely depends on the modality in which the interruption is presented.

Latorella (1996, 1999) has conducted one of few studies on the role of modalities (in her study, the visual and auditory channel) in interruption management. She found that interruptions of an auditory task were acknowledged significantly slower, and were more disruptive, than interruptions to visual tasks. Responses to interruption in crossmodal conditions were significantly slower than in same-modality conditions. And the performance of interrupting tasks was begun significantly slower when the interruption was presented visually, which may be explained by the fact that the long-term availability of the visual reference allows for delays in attending to an interruption.

A recent study in our laboratory also examined the benefits of providing operators with information about the nature of an interruption. In this case, subjects were informed about the urgency and modality of interruption tasks as well as the time remaining to perform those tasks (Ho, Nikolic & Sarter, 2001). 48 subjects performed a simulated air traffic control task. At times, they were asked to perform a secondary task, which consisted of counting the fast patterns in a sequence of fast and slow pulsing patterns that were presented in the visual, auditory, or tactile modality. Subjects were provided with, or could request, information on the urgency and modality of the interruption task and on the time remaining to perform the task.

The findings from this study show that subjects make use of the information on the nature of interruption tasks to manage their primary and interruption tasks more effectively. In particular, knowledge of task modality affected the scheduling of interruption tasks. Subjects who had information about the modality of the interruption tasks delayed the initiation of low priority visual interruption tasks significantly longer than that of the other two types of tasks (see Fig. 3), which were performed more often concurrently with the air traffic control task. Most likely, this served to avoid difficulties resulting from intramodal interference and visual scanning costs. Differences between the experimental groups in terms of task initiation delays were particularly pronounced when subjects were experiencing high workload in their primary ATC task and/or were interrupted frequently.

The results from this study also relate to multimodal timesharing. Effects of resource competition were reflected in significantly worse performance on the visual (as compared to the auditory and tactile) task when this task was time-
shared with the ATC task whereas no significant difference between the visual, auditory, or tactile task was found when these tasks were performed in isolation (see Fig. 4).

These findings confirm the benefits of providing information on the nature of interruption tasks to help operators manage potential interruptions in a more effective manner. Still, one important limitation of the approach used in the above study is that subjects were required to divert their attention away from their primary visual task for a brief period of time to access the visual information on the nature of the interruption task. A more effective way of providing interruption-related information would be to present it via other modalities such as hearing or touch.

By doing so, the three main criteria for supporting preattentive reference would be met: (a) the signal can be picked up in parallel with ongoing lines of reasoning, (b) partial information is presented to help the operator decide whether a shift in attention towards the interrupt signal is warranted, and (c) the evaluation of this information is mentally economical (Woods, 1995). Support for preattentive reference could benefit users in a variety of domains, not just pilots on modern flight decks. For example, problems related to driver distraction associated with the use of cellular car phones (as well as the use of other wireless communication devices) could be reduced by helping drivers make more informed decisions about whether and when to accept an incoming phone call. Knowledge of the identity of the caller and the urgency of the call, for example, could be presented in auditory form to avoid that the driver’s visual attention is diverted from the primary task of driving the car. In combination with the driver’s knowledge of the current driving circumstances (traffic

Fig. 3. Task Initiation Delay as a Function of Task Modality.
volume and complexity, weather and road conditions), this information could enable the driver to manage phone calls more effectively. Similarly, operators in other domains requiring interaction and coordination with other human and machine agents (such as air traffic control, emergency medical care, or process control) could benefit from preattentive reference to help them prioritize and schedule multiple tasks and handle attentional demands in an effective and timely manner.

**CONCLUDING REMARKS**

Breakdowns in human-automation coordination in a variety of complex event-driven domains can be explained, in part, by the almost exclusive reliance in interface design on visual information presentation. This approach fails to exploit the affordances associated with other modalities, such as hearing, touch,
or smell. In this chapter, we have argued that the development and introduction of multimodal interfaces can be a powerful means of supporting coordinative functions such as resource allocation, attention capture and guidance, and interruption management. There is already limited empirical evidence that supports this claim. However, far more research is needed to better understand the affordances and limitations of multimodal systems and improve their viability for real-world environments.

First, more effective and less intrusive technical implementations of delivery systems are needed, especially in the areas of touch and smell. It is likely that progress in this area will be made in the context of the development of VR technologies. In the area of passive multimodal displays, we need to examine the effects of presenting information simultaneously via multiple sensory channels. Most research to date has considered the concurrent use of two modalities only, and most of these studies have focused on visual and auditory feedback but failed to include other senses (for exceptions, see e.g. Spence et al., 1998; McGuirl & Sarter, 2001).

More research is needed also on the context-dependent adaptation and integration of modalities. This will require that we abandon the current focus in research on either system input or output and instead consider the entire cycle of perception and action that constitutes multimodal interaction. In order to create robust adaptive multimodal systems, effective architectures and protocols for interruption management need to be developed. McFarlane (1999) has conducted one of few studies in this area. He examined four possible interruption protocols: immediate, negotiated, mediated, and scheduled. Overall, his research has shown that the choice of protocol depends on the nature of the task and associated performance criteria. In his study, negotiated interruptions worked best for accuracy on continuous tasks but carried the risk that people would not handle interruptions in a timely manner. Immediate interruptions, on the other hand, lead to prompt and effective handling of the interruption at the expense of more errors on the interrupted task.

Very few studies have examined when and how people switch between modalities for different tasks and purposes. One of few exceptions is the work by Oviatt and colleagues on people’s “contrastive functional use of modes” and on the identification of “natural integration patterns that typify people’s combined use of different input modes”. For example, this research has shown that modality switches are often used in an attempt to resolve errors in human-machine interaction (e.g. Oviatt & VanGent, 1996).

In conclusion, multimodal systems are a promising and exciting new means of improving the effectiveness of human-machine interaction and coordination. Recent technological advances have made such systems more feasible but a
wide range of issues still needs to be addressed. Doing so will be critical to avoid that multimodal systems create the kinds of unanticipated difficulties that earlier technologies have introduced to many fields of work.

ACKNOWLEDGMENTS

The preparation of this chapter was supported, in part, by a CAREER award from the National Science Foundation (NSF) (grant No. IIS-9996448; Technical monitor: Ephraim Glinert). I would also like to thank Mark Nikolic for his helpful comments on an earlier version of this manuscript.

REFERENCES


3. OPERATORS’ AUTOMATION USAGE DECISIONS AND THE SOURCES OF MISUSE AND DISUSE

Hall P. Beck, Mary T. Dzindolet and Linda G. Pierce

ABSTRACT

This chapter focuses upon operators’ automation usage decisions (AUDs), choices in which people have the option of relying on automation or employing a manual or less technologically advanced means of control. Misuse, the over-utilization of automation, and disuse, the under-utilization of automation, result from inappropriate AUDs. Three causes of misuse and disuse were identified. Operators may: (1) not recognize that both automated and manual alternatives are available (recognition errors), (2) inaccurately estimate the utilities of the automated and/or manual options (appraisal errors), or (3) knowingly select the alternative with the lowest expected outcome (action errors). A decision-making model was used to organize the literature, identify new areas of inquiry, and examine the effects of a number of variables known to affect AUDs. These
included machine reliability and consistency, performance feedback, trust in automation, self-confidence in manual control, perceived utility, and the automation bias.

OPERATORS’ AUTOMATION USAGE DECISIONS AND THE SOURCES OF MISUSE AND DISUSE

As the new century progresses, the success of human activities will increasingly depend upon how effectively people collaborate with their machine partners. Historically, automation was developed to augment the senses or to enhance muscle power. Automation is now rapidly expanding beyond its traditional role to include the simulation of many human intellectual functions. The productivity of previous generations was largely determined by collective skills, the ability of persons to integrate their activities to achieve a common objective. In the future, efficient individuals must be able to coordinate their actions with those of “intelligent” machines as well as other people.

The ability to simulate a human task does not necessarily mean that a machine instead of a person should perform that activity. Automation will only make for a better and happier way of life if wise decisions are made concerning which tasks or functions should be automated and which humans should retain. While some task allocation decisions are made during the design of the system (Parasuraman, 2000; Sheridan, 2000; Sheridan & Parasuraman, 2000), others are left for operators to determine. This article focuses upon operators’ automation usage decisions (AUDs), situations in which people have the option of performing a task manually or relying on automation.

Automation Usage Decisions

Each day members of technologically advanced societies make dozens of AUDs. One can drive a car or walk to a nearby store, dial a phone number one digit at a time or speed dial, calculate taxes by hand or enter the data into a computer program. Often human-machine teams perform redundant activities. For example, a submariner may conduct a periscope scan of the surface while sonar sweeps the area for other vessels.

When operators choose automated over manual alternatives, they often assume a different set of risks. In a very important paper, Parasuraman and Riley (1997) identified two types of problems that can follow from inappropriate AUDs. Disuse is under utilization, manually performing a task that could best be done by automation. Misuse is over reliance, employing
automation when a manual alternative would have achieved a better end. Knowing whether automated or manual control will yield the more desirable outcome is frequently a very complex proposition. Efficiency in a highly automated society requires discriminating operators, who can determine when to and when not to depend upon machines.

Although misuse and disuse are recently coined terms, their underlying social, emotional, and cognitive processes have been extensively studied in other literatures. A fruitful tactic (e.g. Wickens & Flach, 1988) has been to use decision-making models to guide research in automation usage. Like Wickens and Flach, this paper will take conceptual direction from studies of human information processing. Our purposes are to employ a decision-making model to: (1) organize much of the research relating to AUDs, and (2) generate unexplored hypotheses.

A Few Definitions and the Domain of AUDs

Before proceeding with our analysis of AUDs, some important definitions need to be examined. Automation is “any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by a machine” (Moray, Inagaki & Itoh, 2000, p. 44). It is essentially the simulation of human activity by machines. Usually the move from manual to automated control is not all-or-none. A number of classification systems (e.g. McDaniel, 1988; Riley, 1989; Sheridan & Verplank, 1978; Wickens, 1997) incorporate the idea that automation is most accurately viewed along a continuum.

The vertical axis in Table 1 shows the ten levels of automation originally proposed by Sheridan and Verplank (1978). These categories were designed to represent output or post-decision functions. Manual operation, in which a person carries out an activity without intervention from a machine, is at one end of the continuum. At the higher levels of the scale, the machine becomes increasingly autonomous of human control. The tenth level is full automation; the machine performs all aspects of the task independent of human intervention. Recently, Parasuraman, Sheridan and Wickens (2000) extended this model to include input functions. Their four-stage information-processing model was meant to aid system designers in determining if and to what extent particular functions should be automated.

The first stage, “information acquisition,” includes the sensing and registration of input data. Telescopes, which can automatically be directed to a particular star or planet, are an example of first stage information processing. The second stage called, “information analysis” involves cognitive functions such as working memory and inference. Actuarial equations and other algorithms used
to predict future events are instances of this stage. "Decision and action selection" is the third stage and consists of a comparison of alternatives. Costs and benefits are weighed and the best option is determined. Course-of-action software (National Simulation Center, 1998) that suggests the fastest and safest route for moving troops across the battlespace performs the decision and action selection function. "Action implementation" is the final stage and refers to the response following the decision. Programming a robotic arm to retrieve a satellite is an example of action implementation.

Although Parasuraman et al.'s (2000) model was intended as an aid to system designers, it is also useful for delineating the boundaries of research inquiries. Automation usage decisions can occur at each of the four stages of information processing.

Table 1. Automation Usage Decisions as a Function of Information Processing and Automation Levels.

<table>
<thead>
<tr>
<th>Information Acquisition</th>
<th>Information Analysis</th>
<th>Action Selection</th>
<th>Action Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The computer offers no assistance: human must make all decisions and actions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The computer offers a complete set of decision/action alternatives, or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. narrows the selections down to a few, or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. suggests one alternative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. executes that suggestion if human approves, or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. allows the human a restricted time to veto before automatic execution, or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. executes automatically, then necessarily informs the human, and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. informs the human only if asked, or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. informs the human only if it, the computer, Decides to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. The computer decides everything, acts autonomously, ignoring the human.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table was adopted from Parasuraman, Sheridan and Wickens (2000). Shaded areas represent functions and automation levels requiring automation usage decisions.
processing. Astronomers can manually or automatically direct small telescopes, predictions may be based on subjective opinion or statistical regressions, commanders can ignore or accept the advice of decision aids and planes can be flown manually or by autopilot. Unlike the stages of information processing, some levels of automation preclude the possibility of AUDs. Human-machine teams are found at levels two through nine, but only the second through the sixth levels allow operators to choose between automated and manual control. If Parasuraman et al.’s (2000) model is viewed as a $10 \times 4$ matrix with levels of automation and stages of information processing as factors, our purview crosses the four information processing functions at automation levels two through six (see Table 1).

The current area of inquiry is further restricted in that not all operator failures are misuse or disuse. Misuse and disuse must be differentiated from processing and/or behavioral mistakes that prevent the successful execution of an automated or manually controlled system. For instance, Souhrada (1995) reported that a nurse erroneously read a digital display, increased the delivery rate on an infusion pump and overdosed a patient. An incorrect response may also have caused the sinking of the USS Squalus on May 23, 1939 (Maas, 1999). Although no conclusive evidence was ever found, one hypothesis is that the main induction valve was inadvertently opened after the Squalus submerged. A sailor may have pulled the wrong lever and sank the submarine. In both of these situations, it was not the decision to use automation that was inappropriate. The problems were ones of response, how the automation was applied.

It may already be apparent that the word “manual” is employed differently than in general conversation. Push mowers are the manual alternative to power mowers, but manual does not imply that the operator is whacking grass with his or her hands. Manual and automated control are relative terms, manual referring to the less automated of the two alternatives. The advantage of relative terminology is that the investigator does not need to classify a device according to an absolute metric of technological sophistication. The disadvantage is that what is manual in one comparison may be automated in another. The push mower, which joins a wheel to a blade, is an impressive piece of automation. This technological achievement can be fully appreciated by cutting grass for several hours with a less automated scythe.

Describing the automated alternative as an “aid,” “partner” or “teammate” might also create a terminological snarl. These descriptors emphasize that the machine is conducting an activity that could conceivably be done by a person. The terms also reflect the collective nature of the operator’s interaction with the automated device; the human and computer form a team (Bowers, Oser, Salas & Cannon-Bowers, 1996; Bubb-Lewis & Scerbo, 1997; Scerbo, 1996;
Woods, 1996). By choosing the automated over the manual option, the operator assigns responsibility to the machine for performing a task or carrying out an activity.

A distinction also needs to be made between measures of reliance and performance. Reliance is the tendency to employ automated rather than manual control. Selecting the automated option 80% of the time exhibits greater reliance than selecting the automated option 50% of the time. Avoiding misuse and disuse is often a question of balance. Operators are vulnerable to misuse if they overly rely on automation, exaggerating the benefits of machine over manual control. Conversely, persons, who rely too much on manual control, are susceptible to disuse. The best chance for achieving the equilibrium that will yield appropriate automation use is to understand the AUD process.

*The AUD Model*

In a two-choice situation, in which the operator must choose between automated and manual alternatives, the optimal decision is the one that has the highest probability of attaining the objective. Misuse and disuse are the products of sub-optimal AUDs in that they do not maximize the chance of a successful outcome. An examination of the empirical literature, informal discussions with participants in our experiments and the investigators’ own observations led to the identification of four common causes of inappropriate automation use. Automation usage errors occur because operators: (1) fail to recognize that automated and manual alternatives are available, (2) inaccurately appraise the utilities of the automated or manual options, (3) knowingly select the alternative with the lowest expected outcome, and/or (4) do not emit the responses required for successful performance.

If arranged sequentially, these sources of error closely correspond to the processes that operators often go through in making AUDs. That is, AUDs frequently involve realizing that a task could be done under manual or automated control, some estimation of the likelihood that the alternatives will be successful, the selection of either the manual or automated option and then performing the actions necessary to complete the task. Figure 1 shows the four components of what we refer to as the AUD Model.

Lest the reader conclude that several decades of research have been ignored (e.g. Beach, Chi, Klein, Smith & Vicente, 1997; Klein, 1997; Zsambok & Klein, 1997), we should acknowledge several attributes of the model. First, the AUD Model does not assume that particular operators are or are not consciously aware of why they chose the automated or manual option. Operators often recognize, appraise, and select between automated and manual alternatives.
STEP 1: RECOGNITION OF ALTERNATIVES

Realize that the activity could be performed by automation.

Realize that the activity could be performed manually.

STEP 2: APPRAISAL OF OPTIONS

Estimate the likelihood that automated and manual control will achieve the objective.

Estimate the gains and losses associated with each option.

Rate the alternatives.

STEP 3: ACTION SELECTED

Choose the option with the highest expected outcome.

STEP 4: RESPONDING

Perform the actions necessary to accomplish the objective via automated or manual control.

Fig 1. The Four Components of the Automation Usage Decision Model.
with little or no conscious awareness. Also, we are not suggesting that operators making AUDs invariably go through each of the four steps of the decision-making process. Over the past four years, the AUD Model has evolved as a helpful analytic tool in our laboratories. The components of the model provide a structure for organizing the literature, point to gaps in the research, and assist in discovering why sub-optimal AUDs are made. Each of the four steps of the model, or if you would prefer the four sources of inappropriate automation use, will be discussed in turn.

Recognition of Alternatives
Recognition is included as the initial phase of an AUD because operators sometimes do not realize that an activity could be performed manually or via automation. For example, an investigator was recently composing an article about automation misuse, when the computer server went down, temporarily denying him access to his word processor. For 15 minutes he sat, disgusted at the forced inactivity, before remembering that he used to write with a pencil. Failure to recognize that automated and manual alternatives are available does not insure inefficiency, but it increases the probability of sub-optimal AUDs.

Appraisal of Options
After recognizing that the automated and manual options are available, the operator must appraise each alternative. Appraisal consists of two processes, estimating the likelihood of achieving the objective through automation and manual operation and comparing the two assessments (see Fig. 1). Humans possess a veritable potpourri of biases that may distort appraisals and cause misuse and disuse. Some of these biases are specific to automation usage (e.g. Mosier & Skitka’s (1996) automation bias), while others are part of the individual and social cognitive literatures (e.g. Gollwitzer & Wicklund’s (1985) self-serving biases).

Action Selection
Avoiding recognition and appraisal errors does not guarantee an optimal AUD. People frequently know what is best and yet elect to behave contrary to their own welfare. Choosing an automated or manual alternative that the operator realizes has a lower probability of success will be termed an action error (Nagel, 1988), or an “action bias.” A college professor recently provided an example of this form of bias. Most instructors maintain and compute the grades for large classes on spreadsheets, rather than old-fashioned grade books. Therefore, it was surprising to see a colleague, who is skilled
with spreadsheets, calculating the final marks of a class of 50 from a grade book. Because this professor had just complained about the short time before the grades were due, the bias was particularly striking. When asked why he was using a grade book, he claimed that writing down and calculating marks made him feel closer to his students. The professor’s AUD was not an appraisal bias. He intentionally decided to use the slower and less accurate grade book instead of the faster and more accurate spreadsheet. Like this professor, persons exhibiting action biases are often responding to idiosyncratic goals or objectives. As we will soon see, action biases may also arise from previously established appraisal biases.

Responding
The responding phase is not actually part of an AUD or a cause of misuse or disuse. It is included to acknowledge that once the operator chooses between the manual or automated options, some action by the person and/or machine is usually necessary to carry out the activity. Errors in responding, rather than recognition, appraisal or action biases, are sometimes the cause of accidents and other failures.

The crash of American Airlines Flight 965 near Cali, Columbia in December 1995 is a tragic example of a responding error (Phillips, 1999). In a rush to descend, the pilots programmed the wrong abbreviation into the flight management system. The plane changed its direction and flew into a 9000 ft mountain; 159 of 163 persons on board were killed.

Now equipped with this model, let’s see how the research fits upon it.

FACTORS AFFECTING APPRAISAL AND ACTION SELECTION

Machine Reliability and Consistency

Modern workplaces are caught in a never-ending spiral of technological improvements. Most of this new technology is purchased under the assumption that better hardware will augment productivity. Although better machines often enhance production, potentially valuable automation is sometimes introduced into the workplace with little or no improvement in human-team performance (Baxter, 1989; Bowes, Kenney & Pearson, 1993; Harris, 1994; Schneider & Klein, 1994).

Experiments, assessing the association of machine reliability to performance, have also yielded inconsistent results. While some studies (e.g.
Moray, Inagaki & Itoh, 2000; Riley, 1994) found that reliability increased the tendency to depend on automation, others did not obtain this association. For instance, Parasuraman, Molloy, and Singh (1993) and Singh, Molloy and Parasuraman (1997) required operators to simultaneously perform three flight simulation tasks. Two of the tasks required manual operation, but a third monitoring task was automated. The experimenters manipulated the reliability of the device and its consistency. Participants assigned to the consistent conditions, worked with an automated system that was correct either 87.50% or 56.25% of the time during the session. Operators paired with inconsistent automation, worked with a machine whose accuracy switched between 87.50% and 56.25% at ten-minute intervals. Although overall reliability was not related to reliance, machines that maintained the same level of reliability throughout the session were relied on to a greater degree than were devices whose accuracy varied.

A signal detection analysis, performed by Sorkin and Woods (1985), provides additional evidence that the relationship between machine reliability and performance is quite complex. They calculated that the response criterion that optimized an automated aid’s performance did not maximize the performance of a human-computer team. If the computer is considered analogous to a human partner, Sorkin and Woods’ findings are in accord with results from the team performance literature. Group studies (Bass & Barrett, 1981; Steiner, 1972; Terborg, Castore & DeNinno, 1976) often find modest or no relation between the average skill levels of individuals and overall team performance.

**Machine Reliability and the Soldier Detection Procedure**

Our laboratories have conducted several experiments assessing the effects of machine reliability, using what is called the soldier detection paradigm. In the first study, Dzindolet, Pierce, Beck, Dawe and Anderson (2001) began each trial by briefly showing the participant a slide on a computer monitor. Some slides included a soldier (see Fig. 2 for an example) in various levels of camouflage, while others were only of terrain at Fort Sill, OK. Next, an automated aid, called a “contrast detector,” attempted to determine if the photo contained a human form. The trial concluded when the participant indicated via a mouse click if a soldier was in the photograph. This procedure was designed to give operators the option of taking or ignoring the “advice” of the detector. The primary performance measure was the number of slides that were correctly identified.

Prior to beginning the study, three groups of university students were told that the contrast detector would be correct on 60%, 75% or 90% of the trials,
respectively. A fourth group of participants performed the soldier detection task without an aid. They simply observed the photograph and indicated if they believed that the camouflaged soldier was present or absent in the slide. Inclusion of the alone condition allowed investigators to determine if the aid facilitated or impaired performance. Students then responded to 226 actual trials with the experimenter outside the room.

Analysis of variance yielded no significant differences in the probabilities of errors associated with the four conditions ($M_{\text{without-aid}} = 0.16$, $M_{-60\% \text{aid}} = 0.16$, $M_{-75\% \text{aid}} = 0.13$, $M_{-90\% \text{aid}} = 0.12$). None of the human-machine teams were superior to manual control, nor did the reliability of the aid affect accuracy. The performance of the human-machine teams was also compared to full automation (Sheridan & Verplank’s, 1978, Level 10). Had the investigators assigned full decision-making authority to the aid that was correct 60% of the time, it follows that the probability of an error averaged over trials would be 0.40. Giving participants authority to override the aid that was 60% accurate had a beneficial effect on performance. The probability of an error of operators paired with a detector that was 60% accurate was only 16. Permitting humans to make the final decision also had a facilitative effect when the aid

Fig. 2. Sample Slide Showing Fort Sill Terrain and a Soldier.
was accurate on 75% of the trials. Instead of 0.25, the probability of an error with human input was 0.13. Allowing the human to override the detector with 90% accuracy slightly increased the probability of an error from 0.10 to 0.12. These results indicate that the value of the human to the decision-making process declined as the reliability of the aid increased. Full automation was only a viable alternative with the most accurate contrast detector.

One interpretation of the performance data is that the groups paired with a detector did not differ from manual control, because students disregarded the aid’s recommendations. If the participants were unaffected by the detector, then the aid’s decision should not influence the likelihood of an incorrect response by the operator. That is, the probability of an operator error when the aid was correct ($p(\text{error} \mid \text{aid correct})$) should equal the probability of an error when the aid was incorrect ($p(\text{error} \mid \text{aid incorrect})$). A reliable difference between the probabilities associated with correct and incorrect recommendations would suggest that the detector’s responses influenced the operators’ decisions.

A 3 (Aid Reliability: 60%, 75%, 90%) × 2 (Aid Recommendation: Correct, Incorrect) ANOVA was conducted in which Aid Reliability was a between-subjects and Aid Recommendation was a within-subjects variable. The probability of an error was the dependent measure. The main effect for Aid Recommendation attained statistical significance. The probability of an operator error following an incorrect recommendation by the machine (0.27) was greater than the probability of an error following a correct recommendation (0.13). Thus, operators’ decisions were related to the detector’s recommendations. Aid Reliability did not attain statistical significance, revealing that the extent of reliance did not depend upon the accuracy of the machine. Apparently, operators equipped with automation were affected by the aid’s advice, but their willingness to rely on the machine was not governed by the accuracy of the contrast detector.

Given that the operators were influenced by the automated aid’s decisions, why didn’t the reliability of the automated aid affect overall performance? The most probable interpretation of these data is that reliance on automation had both beneficial and detrimental effects. Furthermore, the positive and negative effects of reliance must have been of approximately equal magnitude. A conceptual replication (Pomranky, Dzindolet, Pierce, Beck & Peterson, 2000) of Dzindolet, Pierce, Beck, Dawe and Anderson’s (2001) study was later performed using a within-subjects, rather than a between-subjects design. Like the initial investigation, Pomranky et al. (2000) did not find that the reliability of the automated aid significantly affected overall performance. These results along with those of other investigations (Parasuraman, Molloy & Singh, 1993;
Singh, Molloy & Parasuraman, 1997) demonstrate that system designers should never assume that more reliable decision aids will necessarily produce better performance by human-machine teams. This is a point made years ago by among others Halpin, Johnson and Thornberry (1973) in their studies on cognitive reliability and human computer interface design.

Prevalence of Appraisal Biases and Action Biases
According to the AUD Model, appraisal and action selection biases are two of the most important causes of misuse and disuse. Although most people would probably accept that operators sometimes incorrectly estimate the relative utilities of the automated and manual options, the significance of action biases is more controversial. If humans are basically rational beings, then only rarely should operators intentionally lower the likelihood of success on a task. Conversely, if operators are primarily motivated by unconscious and non-rational processes, a position exposed so well by Freud (Brill, 1938), then action biases may well be commonplace. An experiment by Beck, Dzindolet and Pierce (2001), parts of which have been reported elsewhere (Hawkins, Dzindolet & Beck, 1999; Moes, Knox, Pierce & Beck, 1999), used the soldier detection paradigm to assess the frequency of appraisal and action biases.

As in Dzindolet, Pierce, Beck, Dawe and Anderson’s (2001) study, the task required undergraduates to determine if a soldier was present on a slide shown for a short duration. A critical procedural change from the earlier investigation was that participants first made their decisions and then saw the contrast detector’s choice. Thus, operators could estimate their accuracy and the detector’s accuracy, but could not use the machine to determine if the soldier was present on a particular trial.

The design was a 2 (Machine Performance: Inferior, Superior) × 2 (Feedback: Present, Absent) in which all variables were between-subjects. The computer program was written so that the contrast detector made approximately twice as many errors (machine inferior) or half as many errors (machine superior) as the participant did. After 200 trials, half of the students in each condition were told how many errors that they and the machine had made (feedback present). The remaining participants received no feedback (feedback absent).

Extra credit was determined by randomly selecting and replaying ten of the previous trials. Students were told that the amount of extra credit that they earned depended solely on how many of these trials were answered correctly. They were then presented with an AUD that furnished the primary dependent measure. Students could either replay ten trials displaying their decisions or ten trials showing the decisions of the detector.
If the students’ goal was to maximize their extra credit, their AUD should depend on whether they were superior or inferior to the detector. Misuse would occur if a student paired with an inferior machine made extra credit contingent upon the computer. Disuse would result if a student working with a superior machine elected self-contingent credit. Because evidence of misuse is more common than disuse in the experimental literature (e.g. Layton, Smith & McCoy, 1994; Mosier & Skitka, 1996), the main effect for the machine performance variable was predicted to be statistically significant. Participants in the machine inferior condition were expected to make more biased decisions than were students in the machine superior condition.

A main effect for feedback was also hypothesized. To make an optimal or rational decision, students in the no feedback arrangements needed to avoid both appraisal and action biases. That is, errors would result if operators incorrectly estimated their accuracy relative to the detector’s accuracy (appraisal bias) or knowingly selected the option that was likely to minimize their extra credit (action bias). Feedback following the 200th trial was expected to greatly reduce appraisal biases, leaving only action biases to produce non-rational AUDs. Because operators given feedback knew their own accuracy as well as the accuracy of the machine, they only needed to select the more accurate option to make a rational AUD and to maximize their opportunity for extra credit. The weakening of appraisal biases by feedback was expected to substantially decrease or eliminate non-rational AUDs. Such a finding would indicate that most of the misuse and/or disuse in this study resulted from appraisal biases and that action biases were a much less common problem.

A logit analysis yielded a statistically significant outcome for the machine performance variable, but not in the predicted direction. Automation misuse did not occur; students working with inferior machines made no biased decisions. However, biases generated widespread disuse of automation. Thirty-one of 36 (86%) students paired with a superior machine made extra credit contingent on their own performance, a decision that lowered the extra credit that they received. No evidence was found suggesting that feedback affected the frequency of biased decisions.

The most surprising finding was that a high percentage of participants displayed action biases. Fifteen of 18 (83%) students, paired with a superior machine and given feedback, made non-rational AUDs. These students were aware that they would obtain more points if they based their extra credit on the machine’s decisions. Nevertheless, they knowingly and willingly lowered their grades to avoid depending on the contrast detector. Extrapolation of these results suggests that action biases are a frequent problem, one that must be
overcome if operators are to use automation effectively in the workplace and on the battlefield.

Feedback as a Means of Counteracting Biases

While action biases have received little attention in the laboratory, an unwillingness of workers to adopt effective technology is frequently cited as an impediment to improving productivity (e.g. Brosnan, 1998; DiBello, 2001; Holbrook, 1998; Ricks & Squeo, 1999; Salas & Klein, 2001). Often organizations attempt to overcome resistance to new technology by intimidating the workforce. The threat is sometimes subtle and is frequently presented as an educational opportunity, but most workers recognize that beneath the soft glove is a fist. At the least, the arrival of new automation specifies the will of the organization and delineates the path to advancement. Although aversive social pressures may elicit compliance, it seemed to us that there are more effective and certainly more humane ways to attenuate action biases.

Our laboratories’ initial efforts (Beck, Dzindolet, Pierce, Poole & McDowell, 2000) to alleviate action biases were guided by an interpretation of Beck et al.’s (2001) results and the research on belief perseverance. The main finding (e.g. DiFonzo, 1994; Ross & Anderson, 1982; Ross & Lepper, 1980) in the belief perseverance literature is that false ideas frequently continue to influence attitudes after they have been discredited. For example, Anderson, Lepper, and Ross (1980) induced the belief that risk-prone people make better firefighters than more cautious persons. After this idea was established, participants wrote reasons for why high-risk takers are superior firefighters. Later the investigators revealed that the original belief, that risky people are more effective firefighters, was false. Despite the fact that participants now realized that their initial opinion depended on bogus information, many maintained the notion that high-risk takers make excellent firemen.

Belief perseverance solidifies many social attitudes, including prejudice. For instance, some children are exposed to false information that leads them to develop justifications for their dislike of other ethnic groups. As adults, many will realize that their bias against other groups was built on untrue characterizations. Though the falsity of prior beliefs may be acknowledged, these individuals frequently retain a tendency to expect inferior performance or “inappropriate” behavior from members of minority groups. Biases often beget biases that survive the death of the parent belief. If the effects of untrue beliefs can linger, maintaining prejudice against people, then it seems likely that belief perseverance can sustain a similar prejudicial response against a machine that simulates human thoughts and behaviors.
False opinions, weak disconfirmations, and supportive rationales are the critical components of belief perseverance. Discussions with Beck et al.’s (2001) participants after the training trials revealed that these three elements were inadvertently generated by the experimental procedure. Many operators paired with a superior machine incorrectly assumed that they were more accurate than the detector. Furthermore, most participants had little or no difficulty furnishing rationales underpinning their sub-optimal AUDs. Although feedback delivered after the training trials insured that the operators realized that they were less accurate than the detector, this disconfirmation was too weak to influence their credit decisions. In the language of the AUD Model, the effects of previously established appraisal biases persisted, producing the action biases recorded by Beck et al. (2001).

Beck, Dzindolet, Pierce, Poole and McDowell (2000) sought to promote rational AUDs through three feedback manipulations. Each type of feedback was designed to combat a different component of belief perseverance. The first sought to prevent the establishment of false ideas, the second to disconfirm any erroneous beliefs that were formed, and the third to challenge the rationales supporting automation disuse.

The development of the procedure to thwart the establishment of false beliefs was based upon the operators’ descriptions of their AUDs. Many of Beck et al.’s (2001) participants reported trials in which they saw the soldier and yet the machine found no soldier in the photograph. They were positive that the detector had erred. In contrast, the students were never certain that they had made a mistake. Even if their decision was a complete guess, there was at least a 50% chance that they would be right on a particular trial. Operators may have established a false belief in their superiority because the detector’s errors were more blatant than their mistakes.

If this interpretation is correct, then making the operator’s and the detector’s errors equally obvious should reduce the likelihood of false beliefs and decrease non-rational AUDs. Beck et al. (2000) sought to make human and machine misjudgments equally apparent through a manipulation called individual trial feedback. After the operator and detector responded, participants receiving individual trial feedback were told if the soldier was present or absent in the preceding photograph.

Individual trial feedback might decrease the probability of false beliefs forming, but it would be unrealistic to expect that some participants would not continue to incorrectly assess their own performance relative to the detector. Memory is highly selective and not all mistakes have an equal influence on future judgments. For example, when recalling their experiences during the training session, operators may have vivid memories of the detector’s errors,
but less prominent recollections of their own misidentifications. A tendency to weight the detector’s mistakes more heavily than one’s own errors could cause operators to incorrectly estimate that they made fewer total errors than the detector. False beliefs, once established, must be disconfirmed if misuse and disuse are to be avoided.

The feedback manipulation used by Beck et al. (2001), in which participants were told the total errors that they and the detector had made, was a disconfirmation procedure. Although Beck et al. (2001) did not find that this information affected AUDs, similar performance feedback has been shown to be an important factor regulating the decisions and behaviors of individuals (e.g. Van Houten, Nau & Marini, 1980) and teams (e.g. Dyer, 1984; Klaus & Glaser, 1970). This manipulation was referred to as cumulative feedback by Beck et al. (2000).

A third means of changing false beliefs is to attack their underlying rationales. Often rationales are directly challenged by counter arguments. A more indirect approach is to cause people to doubt their reasoning by making their thoughts and conclusions appear odd or unconventional. As Cialdini (2001) summarized, “We view a behavior as correct in a given situation to the degree that we see others performing it” (p. 100). Beck et al. (2000) brought the principle of social proof to bear through an intervention called prior results feedback. Operators given prior results feedback were told how successful participants in previous studies had maximized their credit. Operators paired with a superior machine were informed that persons in earlier experiments who made credit contingent on the detector usually obtained greater rewards than persons who made credit contingent on their own performance. This feedback manipulation was intended to make students inclined towards manual control question the wisdom of their AUD.

The design was a 2 (Individual Trial: Present, Absent) × 2 (Cumulative: Present, Absent) × 2 (Prior Results: Present, Absent) between-subjects in which the three types of feedback served as independent variables. The basic procedure was similar to that used by Beck et al. (2001) except that all operators worked with a superior machine. Over the course of the training trials, the detector made approximately half as many errors as the operator did. On each trial, a photograph was displayed, then the operator responded, indicating if the soldier was present. Lastly, the contrast detector reported the results of its scan.

Students receiving individual trial feedback were told if the soldier was present or absent immediately after they and the detector had made their decisions. Cumulative and prior results feedback were presented via message boxes after the 200th training trial. Once again, extra credit was based on the outcomes of ten trials, which were replayed from the preceding 200. The
dependent variable was the operator’s choice to make credit contingent on their own (non-rational AUD) or the machine’s (rational AUD) performance.

The only main effect to attain statistical significance was cumulative feedback. Providing students feedback of the total number of errors that they and the machine made decreased non-rational decisions. The Individual Trial × Prior Results interaction was also statistically significant. Fewer non-rational decisions were made when operators were given both individual trial and prior results feedback than in the other three combinations of these variables. None of the remaining interactions was statistically significant.

An additional analysis indicated that the more types of feedback that the operator received the less likely they were to make credit contingent on their own performance. Eighty-four percent of the operators given no feedback made non-rational decisions compared to 44% of the operators administered three types of feedback. This finding suggests that operator-training programs would improve their effectiveness by including a combination of individual trial, cumulative, and prior results feedback.

The success of the feedback manipulations in reducing disuse should not obscure the magnitude of the problem caused by action biases. Almost half of the operators receiving three forms of feedback still made non-rational AUDs. There can be no question that these students knew that opting for self-contingent credit would lower their grades. Nevertheless, the strength and resiliency of the action biases was such that they made sub-optimal decisions. Building a firewall to protect against non-rational AUDs will require powerful interventions to counteract the effects of action, as well as appraisal biases. Beck et al.’s (2000) experiment provided a first step in that direction, but stronger procedures must be devised if the influence of action biases on non-rational AUDs is to be controlled and high levels of productivity attained.

**Familiarity, Expertise and Attitudes Towards Automation**

The attitudes and the beliefs that the operator holds are another set of variables that may influence the appraisal and/or action selection processes. Not surprisingly, surveys of attitudes about automation and/or new technologies usually show large individual differences (Chambers, Jarnecke & Adams, 1976; Goldman, Platt & Kaplan, 1973; Tenney, Rogers & Pew, 1998; Weil & Rosen, 1995). In 1973, Halpin, Johnson, and Thornberry cited a national survey that found that while most people believed computers would improve their lives, about half of the sample saw computers as dehumanizing and prone to errors. More recent assessments have continued to find large differences among
respondents in their reactions to technological change. For example, McClumpha and James (1994) and Singh, Molloy and Parasuraman (1993) reported that while some pilots had very favorable views other had negative opinions of cockpit automation.

To some extent, people’s attitudes toward technology can be attributed to their familiarity with these devices. Usually, people who are familiar with automation and/or computers tend to have favorable attitudes (Lee, 1991; Lerch & Prietula, 1989; McQuarrie & Iwamoto, 1990). A study by Bauhs and Cooke (1994) yielded a discrepant finding. Respondents who knew most about how computers worked viewed them as less credible.

Expertise is another variable that is predictive of how people evaluate new technology. Most studies (e.g. Kantowitz, Hanowski & Kantowitz, 1997; Lerch & Prietula, 1989) find that experts in an area hold more negative opinions than nonexperts. There is some evidence (Sheridan, Vamos & Aida, 1983) that experts are inclined to reject potentially useful information from a computer, because they have less need for the information that these instruments provide. The research on expertise implies that the greatest resistance to decision aids in the workplace may come from those individuals who are most skilled in their professions. For instance, a novice commander may eagerly embrace the use of course-of-action software to plan troop movements across the battlefield while a more experienced commander may belittle the suggestion that a computer program might assist in this task.

**Trust in Automation and Self-Confidence in Manual Control**

Trust in automation or in one’s own ability to perform a task is probably one of the most important determinates of the outcome of appraisals and action selection. Trust has been a well-researched construct, yielding an impressive series of investigations (Cohen, Parasuraman & Freeman, 1998; Lee & Moray, 1992, 1994; Moray, Inagaki & Itoh, 2000; Mosier & Skitka, 1996; Singh, Molloy & Parasuraman, 1993; Tan & Lewandowsky, 1996). Muir (1987, 1994) derived a theoretical account of trust in automation by extrapolating from human trust literature. She proposed that automation will be seen as trustworthy if it is predictable, dependable, and inspires faith that it will behave as expected in unknown situations. According to her analysis, trust is gained in the areas of persistence, technical competence, and fiduciary responsibility.

Much of the interest in trust is due to its hypothesized association with AUDs. It seems logical that a lack of trust in a superior aid may be a precursor of disuse while overly trusting an automated aid can cause misuse. Parasuraman and Riley (1997) described several real-world incidences in
which disastrous results occurred because people ignored warning signals that they saw as untrustworthy. Disregarding automated alarm systems that have previously signaled in error has been dubbed the cry wolf effect (Bliss, 1997; Bliss, Dunn & Fuller, 1995). Publicizing the alarm’s trustworthiness was one strategy that proved effective in reducing the cry wolf phenomenon (Bliss et al., 1995).

In an unpublished study, described by Lee and Moray (1992), Muir found that reliance on an automated device to control a simulated orange juice pasteurization plant was positively correlated with trust in the aid. Lee and Moray (1992, 1994) expanded upon Muir’s work, using a more complex version of the orange juice plant simulation. The plant included three subsystems, each of which could be operated manually or automatically. Participants could allocate tasks any way they wished and could change their allocations easily. As part of the experiment, one of the subsystems failed periodically, whether controlled automatically or manually. At the end of each session, participants completed a questionnaire concerning their trust in their automation and self-confidence in performing the tasks manually.

Results indicated strong individual differences in automation use; some participants preferred manual control and others automation. Once the operators assigned a subsystem to automated or manual control, they seldom changed the type of control. Even failures of manual and automated systems were unlikely to alter their AUDs. The most important outcome of Lee and Moray’s (1992, 1994) experiment was that the original task allocation decision could best be predicted by taking the operators’ self-confidence in manual control as well as their trust in automation into account. This finding reveals that operators are engaging in an appraisal process that includes assessment of both the automated and manual options.

Operators’ self-confidence in their abilities to manually perform is related to Bandura’s (1986) concept of self-efficacy. Self-efficacy “refers to beliefs in one’s capacities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 3). The concept of self-efficacy is associated with a number of other constructs studied by social psychologists, including optimism (Seligman, 1991), locus of control (Rotter, 1966), learned helplessness (Abramson, Seligman & Teasdale, 1978), and self-esteem (Jones, 1990). Among other positive attributes, people with strong feelings of self-efficacy are less anxious and depressed (Litt, Nye & Shafer, 1995; Oatley & Bolton, 1985), more persistent in pursuing career objectives (Betz & Hackett, 1986), healthier (Goldman & Harlow, 1993; Maddux, 1991), and more academically successful (Bandura & Schunk, 1981; Beck & Davidson, in press). Individuals’ levels of self-efficacy tend to be situation specific. A person may
have a strong belief in his or her self-efficacy in preparing for an athletic event and a lack of self-efficacy concerning an upcoming music recital.

It is likely that self-efficacy frequently influences AUDs. For instance, operators may become overly dependent on automation if they underestimate their ability to achieve an objective by manual control. Low self-efficacy may have been responsible for a small plane crash described by Parasuraman and Riley (1997). It appears that the pilot lacked confidence in his ability to manually control the aircraft during a snowstorm. Relying excessively on the automatic pilot, he failed to monitor his airspeed and crashed short of the runway. Low self-efficacy can also lead to avoidance of automation. People will not use an automated device if they feel that they cannot perform the responses necessary for its successful operation. A highly unrealistic lack of self-efficacy in regards to automation is characteristic of technophobia.

Lee and Moray’s (1992, 1994) work not only indicates that operators estimated the utilities of the automated and manual options; it suggests that they rated the alternatives. Generally, when trust in the automation was higher than self-confidence, the participant allocated the task to the automated system. However, when trust in the automation was less than self-confidence in manual control, the participant opted for manual control. Dzindolet, Pierce, Dawe, Peterson, and Beck (2000) and Dzindolet, Beck, Pierce, and Dawe (2001) referred to the comparison of alternatives as perceived utility.

The Relationship Between Automation Attitudes and AUDs

While some investigators, (Lee & Moray, 1992, 1994) found that attitudes towards automation were reflected in operators’ AUDs, others (Riley, 1994; Riley, 1996; Singh, Molloy & Parasuraman, 1993) obtained no statistically significant association. One reason for these inconsistent findings is that AUDs are predicated on variables in addition to trust in automation. If we use the social psychological literature as our guide, then it is likely that measures of trust in automation, self-confidence in manual control or other self-report indices will often be weakly correlated with AUDs. Wicker (1969) reviewed several dozen experiments and found low correlations between self-professed attitudes and behavior. For example, students’ views of cheating were weakly associated with actual cheating and attitudes about the value of church poorly predicted attendance on a given Sunday. These modest relationships do not mean that attitudes are never good predictors of behaviors; some investigators have observed very strong correlations (Kraus, 1995). The strength of association appears to depend upon the saliency of the attitude (Carver & Scheier, 1981; Diener & Wallbom, 1976; Snyder & Swann, 1976), its specificity to behavior (Ajzen & Fishbein, 1977; Kim & Hunter, 1993; Morrison, 1989) and
the degree that other social influences are attenuated (Jones & Sigall, 1971; Triandis, 1982). It is probable that these same factors govern the correlations of trust and other attitudes with AUDs. For example, asking participants in the soldier detection paradigm to describe their opinions of an automation device should augment the saliency of that attitude, strengthening its association with subsequent AUDs.

**Biases that Influence Appraisal**

The AUD Model indicates that an optimal decision-maker must accurately estimate the utility of the automated and manual options. If the operator’s trust in automation and self-confidence in manual control do not parallel the actual utilities of the alternatives, then a sub-optimal AUD is likely. While unrealistic trust and self-confidence can cause non-rational AUDs, they are certainly not the only factors that can distort the operator’s judgment and produce misuse or disuse. Two of these, automation bias (Mosier & Skitka, 1996) and polarization bias (Dzindolet, Pierce, Beck & Dawe, 2001), are the products of automation research. The others presented here are not specific to machines or operators; they were drawn from the individual decision-making (e.g. Tversky & Kahneman, 1973) and the social cognition (e.g. Gollwitzer & Wicklund, 1985) literatures. Although the impact of these biases on AUDs has yet to be ascertained, their effects on other types of decision-making are well established. On this basis, we may extrapolate the effects of self-serving biases, illusions of control, illusionary correlations, and the availability heuristic to the appraisal of automated and manual alternatives.

**Automation Bias.** Mosier and Skitka (1996) coined the term “automation bias” to refer to, “the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing” (p. 205). Rather than going through the cognitive effort of gathering and processing information, the data supplied by the automated systems may be used to save energy. Layton, Smith, and McCoy (1994) obtained results that may be interpreted as indicating automation bias. They found that providing pilots with an automated aid’s poor en-route plan reduced the likelihood that they would generate alternative flight plans on their own. Although automation bias may cause instances of misuse, such as in the Layton et al. study, it cannot explain automation disuse.

Presumably, automation bias may occur in various degrees. In its most extreme form, the decision reached by the automated aid is immediately adopted. In a more moderate extreme form, the decision reached by the aid may be given an inappropriately large role in the human’s decision-making process. Although automation bias is restricted to decisions made by machines, it is conceptually
similar to situations in which the views of one person serve as a heuristic replacement for thinking by another.

**Polarization Bias.** Most of us have noticed that our fellows deviate from perfection on even simple and highly practiced tasks. For instance, although most adults can add and subtract reasonably well, no one is immune to basic arithmetic mistakes. Arithmetic and many other skills vary along a continuum. For most persons, the probability of an error is greater than “0” and less than “1.” In contrast, automated devices often work perfectly or not at all. If the numbers are entered correctly, a calculator will give the right answer with each press of the “=” key or the wrong answer every time. Calculators tend to be either completely functional or worthless. These type of experiences have taught us that a human who makes a mistake on one problem may be correct on the next, but that once a machine makes an error, it is broken and cannot be trusted until it is repaired.

The notion that human performance is continuous and machine performance is dichotomous holds true in most situations. However, application of this rule produces appraisal errors when it is used to estimate the performance of decision aids and other devices that make probabilistic judgments. Like a person, a decision aid may make an error in one instance and be correct the next. We refer to appraisal errors that result because the operator incorrectly assumes that a decision aid or other such instrument will always or never work as polarization biases. A polarization bias is hypothesized to push people to have unrealistic and extremely favorable (perfection bias) or extremely unfavorable (rejection bias) views of decision aids.

One of the authors of this chapter had an experience with a decision aid that may reflect a polarization bias. The aid was based on an actuarial equation (Beck, Palmer, Lindau & Carpenter, 1997) designed to assist prison officials determine whether to promote violent criminals from maximum to minimum security. This is an important decision, in part, because promotion is often the first step towards early release. The aid was simple to use and required little effort. Prisoner officials entered the offender’s data into the program and the computer calculated the likelihood of promotion.

The value of the aid should have been evaluated by assessing its accuracy and contribution to the overall decision making process. Instead, some prison officials argued that it was useless, because there was no guarantee that it would always make the best promotion recommendation. For them, an error was proof that the equation was of no utility. Of course, these individuals were fully aware that no human, relying on subjective judgment would invariably be right when making these difficult decisions.
**Self-serving biases.** If we may generalize from the social cognition literature, it is likely that biases related to the operator’s self-image are a frequent cause of sub-optimal AUDs. Often these biases are self-serving, providing protection against anxiety and depression (Snyder & Higgins, 1988). For instance, most people over rate their contribution to a group (Gollwitzer & Wicklund, 1985), inflate their role in gaining favorable outcomes (Whitley & Frieze, 1985), express excessive confidence in negotiations (Neale & Bazerman, 1985), overestimate the tasks that they can complete in an allotted time (Buehler, Griffin, & Ross, 1994) and are overconfident in their ability to forecast future events (Fischhoff & MacGregor, 1982).

If operators hold strong self-serving biases, it is likely that they will overestimate their capacity for performing a task manually and be vulnerable to automation disuse. Pomranky, Dzindolet and Peterson (2001) found that, without feedback, participants erroneously estimated that their performances were better than that of their automated aids, when in fact, the automated aids had made half as many errors as the participants. Operators who did receive feedback (either continuously or at the end of 200 trials) more accurately perceived their performances relative to that of the automated aid.

**Illusion of Control.** Persons having an illusion of control believe that they direct chance events (Langer, 1983). A frequently cited study by Langer (1977) found that persons who picked a lottery number demanded four times the money to sell their ticket as persons assigned a lottery number. Similarly, people prefer to predict the roll of a die instead of guessing its outcome after the toss (Friedland, Keinan & Regev, 1992). If operators feel that they govern chance events, then it is likely that they will exhibit a preference for manual over automated control.

**Illusionary Correlations.** Preconceived notions may produce illusionary correlations (Chapman & Chapman, 1969; Crocker, 1981; Hamilton & Sherman, 1989; McFarland, Ross & DeCourville, 1989; Trolier & Hamilton, 1986) causing people to see relationships between events where none exist. In an early demonstration of illusionary correlations, Ward and Jenkins (1965) gave participants the results of a fictitious cloud-seeding experiment. The investigators arranged the data so that there was no actual association between seeding and the likelihood that it rained. Nevertheless, people who thought that cloud-seeding was an effective procedure felt that the outcome of the experiment validated their views. The research on illusionary correlations suggests that operators forming strong feelings about the value of manual versus automated control will selectively evaluate their experiences. What they will find is
probably just what they expect to find, evidence supporting their AUD. Illusionary correlations may be one reason that some investigators (Lee & Moray, 1992, 1994) find that failure experiences do little to alter operator’s preferences for manual or automated control.

**Availability Heuristic.** Availability is a well-known heuristic in which people judge the frequency of events by their ready access in memory. For example, there are three times as many English words with “k” in the third letter as “k” in the first letter. Still, most people guess that “k” is more often found at the beginning of a word. Tversky and Kahneman (1973, 1974) contend that this error occurs because it is easier to generate words with “k” in the first than the third position. The easy availability of images may explain why many people overestimate certain risks (Allison, McQueen & Schaerfl, 1992), such as the likelihood of airline crashes. It is likely that some experiences that operators have with manual and automated control will be more vivid and readily available to memory than others. As a result, AUDs may be based on the outcome of a few easily recalled events. When these available memories are not representative of the operators’ experiences, non-rational AUDs are the probable result.

**Human-Machine Differences in the Formation of Appraisal and Action Biases**

Although AUDs are affected by many of the same processes that influence collaboration with humans, working with a machine and a person are not identical experiences. A potentially important difference is that persons may inaccurately estimate the value of a decision aid because it is a machine. A recent experiment by Dzindolet, Pierce, Beck and Dawe (2001, Study 1) contrasted the performance expectations of operators paired with human and automated partners at the start of the soldier detection task. Half the participants were told that on each trial, they would decide if the soldier were present and then be given the decision reached by the contrast detector. Other operators were instructed that following their response, they would see the decision of the prior participant.

Before beginning the task, operators estimated the number of errors their partner would make in the upcoming 200 trials. Operators paired with the detector estimated an average of only 14.19 errors (i.e. correct nearly 93% of the time). When asked to estimate the number of errors the human aid would make, participants predicted, on average, 46.17 errors (i.e. correct only 77% of...
the time – only 27% better than chance!). Thus, operators began the task expect-
ing much better performance from the automated than from the human partner.
Comparing these results to those of a subsequent experiment suggests that
performance expectations changed as operators gained experience with the
soldier detection task. While Dzindolet, Pierce, Beck and Dawe’s (2001)
first investigation assessed expectations before starting the task, their second
study measured performance expectancies after 200 trials. When operators
were questioned after practice with the soldier detection task, there were no
differences in their expectations of human and machine partners.
Dzindolet, Pierce, Beck and Dawe’s (2001) original objectives did not include
recording changes in performance expectations as a result of experience.
Therefore, different samples and dependent variables were used in the two
experiments, conditions that compel caution in comparing across studies.
Despite these caveats, we feel confident in making two generalizations from
these findings. First, operators’ pre-experimental experiences led them to
anticipate very different levels of performance from human and machine
partners. Secondly, there is little reason to doubt that initial performance
expectations are often malleable and can change with practice. Initial
differences in performance expectations disappeared as operators gained
experience with their human and machine partners.
Working in tandem with a machine appears to have much in common with
entering a foreign culture. At first, differences are the figure; the people seem
distinct and unlike you. However, the longer you remain in that culture the
more similarities emerge from the background. Many travelers come to believe
that despite variations on the themes of life, people everywhere are basically
alike. Whether, as we become more used to living with intelligent machines,
some operators take the ultimate step and conclude that working with a machines
is just like working with a person remains to be seen.

A TALE OF THREE APPRAISAL PROCESSES

The AUD Model presented in this chapter suggests that one of three appraisal
processes could occur when operators confront an AUD. First, the operator may
assess only the utility of the automated option, never recognizing that manual
control is an alternative. If the appraisal is sufficiently favorable to exceed thresh-
old, the operator will perform the activity using automation. This decision rule
can be summarized as, “if automation seems good enough, then use it.”
A comparable decision strategy exists for manual control, in which operators
disregard the possibility of automation. Operators often conduct tasks manually,
if their personal appraisal reflects sufficient confidence in manual control.
The AUD Model also holds that there are situations in which operators compare the utilities of the automated and manual options. If perceived utility (Dzindolet, Beck, Pierce & Dawe, 2001; Dzindolet, Pierce, Dawe, Peterson & Beck, 2000) corresponds to actual utility, then the best option will be identified. Mismatches of perceived and actual utilities will likely result in misuse or disuse of automation.

Evaluating one, rather than both the automated and manual options saves time and energy. Although a single assessment strategy may produce satisfactory outcomes for most activities, it will lead to misuse or disuse, if the non-evaluated alternative is the superior option. It is very important for investigators to determine when operators will appraise only one of two options and when they will appraise the utility of both the automated and manual alternatives.

Given that appraising one option takes less cognitive energy than appraising two options, it is reasonable to consider that effort may affect the operator’s assessment strategy. Social psychologists have extensively studied the conditions under which people are likely to save and exert effort. Much of this research is applicable to AUDs.

One condition that often reduces exerted effort is collaboration. When working alone, people pull a rope harder (Ingham, Levinger, Graves & Peckham, 1974), clap and shout louder (Harkins, Latané & Williams, 1980; Latané, Williams & Harkins, 1979), and make fewer errors while performing particular cognitive tasks (Jackson & Williams, 1985) than when they are working in a group in which performance is pooled. This finding has been dubbed social loafing because the reduced performance of the individuals is presumed to reflect their lowered motivation when in groups (Latané et al., 1979). Drawing from this literature, Mosier and Skitka (1996) proposed that the responsibility for the human-computer team’s outcome is diffused between the automated aid and the human operator. Providing operators the opportunity to rely on an automated aid’s contributions is hypothesized to decrease their motivation.

Shepperd’s Expectancy-Value Theory (1993, 1998; Shepperd & Taylor, 1999) successfully accounts for much of the social loafing literature and can be applied to the study of AUDs. According to this theory, motivation is predicted by a function of three factors: expectancy, instrumentality, and outcome value. The first factor, expectancy, is the extent to which members feel that their efforts are necessary for the group to succeed. When members regard their contributions as dispensable, they are less willing to work hard. It is likely that operators will feel that they are dispensable and experience motivation loss if they perceive that automation will accomplish the objective without their input.
Instrumentality is the belief of group members that a successful performance will lead to a positive overall outcome. Members who feel the outcome is not contingent on the group’s performance are less likely to work hard (Shepperd, 1998; Shepperd & Taylor, 1999). This implies that operators who believe their human-computer team’s performance is unlikely to affect the outcome will not exert much effort.

The third factor, outcome value, is the difference between the importance of the outcome and the costs associated with working hard. Increasing the costs or minimizing the importance of the reward will lead members to put forth less effort. More effort will be extended toward tasks that lead to valuable outcomes without requiring much cost.

When less effort is expended due to low expectancy, instrumentality, or outcome value, then we may hypothesize that options are less likely to be fully appraised. Presumably, low motivation resulting in an incomplete appraisal, could augment the manual or the automated alternative. If manual is the operator’s preference, the failure to appraise automation should encourage the continued application of manual control. Likewise, human operators who consider only the automated system, will be unlikely to change to manual operation, even following automation malfunctions.

It is important to recognize that automation is not always a viable escape for individuals seeking to minimize their efforts. Rather than decreasing effort, the introduction of automation often changes the operator from that of an active participant to that of a passive monitor. Desmond, Hancock, and Monette (1998) found that monitoring an automated system that drove an automobile was equally fatigue inducing as actually driving the automobile (manual operation). Automation is sometimes more of disappointment than a refuge for persons hoping to save energy or to slow the pace of their hectic lives.

A particularly interesting issue is the effect of motivation (and thus, arousal level) on the human-computer team’s performance. A basic tenet of Hull’s (1943) theory is that generalized drive and habit strength interact multiplicatively, thereby increasing the probability of the dominant or most frequent response. Application of Hull’s views to AUDs indicates that arousal increments could augment the tendency to rely on either automated or manual control, depending on which choice is dominant. If the dominant response is also the best response in a given situation, high levels of arousal should increase rational AUDs. Conversely, when the best response is not dominant, elevations of arousal should augment misuse or disuse. If rational AUDs become more common or dominant with training or practice, Hull’s theory makes a critical and counterintuitive prediction. Although novices may be especially likely to make non-rational AUDs in highly arousing situations,
well-trained operators may actually respond more rationally when highly stressed.

*Will All the Non-rational Operators Please Stand?*

This request would produce a modest response, even if Diogenes filled the audience with honest people. The reason that most persons would not leap to their feet can be ascertained by a consideration of objectives. Our examination of AUDs has thus far assumed that the operator’s goal is to obtain the best, or at least a satisfactory, performance. Although this assumption is often warranted, other unrecognized objectives may be regulating the operator’s decisions. An example from the Gulf War may prove helpful.

Many soldiers, who took part in that conflict, were instructed in the use of systems in which automation was the first option and manual control was intended as a backup. Training exercises indicated that a desirable balance of automated and manual control had been attained. Unless circumstances definitely called for manual operation, most soldiers depended on automation. Despite their apparent acceptance of these machines, some soldiers turned off their automated instruments and relied on manual control during battle. They did not fight as they were trained.

Several post hoc interpretations may be put forth to explain why soldiers were unwilling to use automated instruments in battle. Not using automation may have resulted from the stress, excitement, and chaos of war. Poorly designed equipment, inadequate training, and the personal defects of particular soldiers may also have contributed to a reluctance to use automation. Though these interpretations are tenable, they overlook that the soldiers’ objectives in training were probably different from their objectives in battle. The main goal of many soldiers during training is to please their superiors. They recognize that the U.S. military is becoming more technologically advanced and that the twenty-first century Army will increasingly depend upon automation. While some soldiers sincerely valued the new technology, others probably feigned enthusiasm to impress their superiors. Battle changes priorities. Questions of life and death become paramount and impression formation becomes secondary. The soldiers’ true assessment of the automated instruments was revealed by a change in objectives.

Knowledge of an objective is often all that is necessary to discover why an operator selected an automated or a manual alternative. Whether an AUD is classified as rational, non-rational, misuse or disuse often depends upon whom is applying the label. As experimenters we must never forget that the majority of people live under the happy premise that they are basically reasonable and
that most of their decisions, including their AUDs, are solid. Most of us see ourselves as logical, even if we may harbor some doubts concerning our neighbors. Automation usage decisions cannot be fully understood if the only viewpoint examined is that of an investigator observing an operator. We also need studies that see AUDs from the operator’s perspective.

**A SUMMATION AND SOME AREAS OF FUTURE INQUIRY**

This chapter examined AUDs, situations in which operators have the opportunity to perform a task or activity using automation or manual control. Though many AUDs go unnoticed, they nevertheless have a profound impact on productivity in the workplace, life in the home, and safety on land and in the air. An increasingly significant question that researchers must address is how operators should best work with machines that simulate human intellectual functions. Within the next several decades, decision aids will be our omnipresent companions providing advice on everything from upgrading computers to when it is an advantageous time to have children. One minor effect of the proliferation of decision aids will be that investigators will no longer need to justify the importance of studying AUDs.

Making the best AUDs is often a question of balance. Too strong a tendency to employ automation may result in misuse; too great a tendency to depend on manual control can produce disuse. For many systems, the most effective AUD will depend upon the situation, the characteristics of the machine, and the attributes of the operator. Although the pace of research is accelerating, experimenters are just beginning to assess the variables affecting AUDs.

To date, investigations of AUDs have primarily focused upon machine reliability, machine consistency, trust in automation, self-confidence in manual control, perceived utility, performance feedback, and attitudes towards automation. Table 2 provides a summary of the AUDs literature presented in this chapter. Taken together, these studies led us to develop a conceptual structure that was the focus of this chapter. This framework, which we refer to as the AUD Model, postulates that misuse and disuse arise from three sources. Sub-optimal AUDs are likely to occur when operators: (1) fail recognize that both automated and manual alternatives are available (recognition errors), (2) inaccurately assess the utilities of the automated or manual options (appraisal errors), or (3) intentionally select the alternative with the lowest expected outcome (action errors). The AUD Model has provided our laboratories with a
<table>
<thead>
<tr>
<th>Variables</th>
<th>Findings</th>
<th>Pertinent Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability:</strong> The accuracy of the machine or the likelihood that an objective can be achieved by automation.</td>
<td>Increasing machine reliability, sometimes but not always, augments reliance on automation. One implication of these studies is that system designers should not assume that providing operators with better automation will necessarily improve the performance of human-machine teams.</td>
<td>Dzindolet, Pierce, Beck, Dawe &amp; Anderson, 2001; Moray, Inagaki &amp; Itoh, 2000; Parasuraman, Molloy &amp; Singh, 1993; Pomranky, Dzindolet, Pierce, Beck &amp; Peterson, 2000; Riley, 1994; Singh, Molloy &amp; Parasuraman, 1997.</td>
</tr>
<tr>
<td><strong>Consistency:</strong> The extent that machine reliability does not change during the session.</td>
<td>Machines maintaining a consistent level of reliability are relied upon more than machines whose levels of reliability vary.</td>
<td>Parasuraman, Molloy &amp; Singh, 1993; Singh, Molloy &amp; Parasuraman, 1997.</td>
</tr>
<tr>
<td><strong>Bias:</strong> A cognitive inclination that leads the operator to over or under value automated or manual control.</td>
<td>Research with the soldier detection paradigm finds that disuse is a common problem. Many students knowingly reduced their probability of earning rewards to avoid relying on an automated aid. Appraisal errors cannot fully account for these non-rational AUDs; action biases were also prevalent. Building a firewall to fully protect against sub-optimal AUDs will require interventions that counteract the effects of action, as well as appraisal biases.</td>
<td>Beck, Dzindolet &amp; Pierce, 2001; Beck, Dzindolet, Pierce, Poole &amp; McDowell, 2000; Dzindolet, Pierce, Beck &amp; Dawe, in press.</td>
</tr>
<tr>
<td><strong>Feedback:</strong> Providing operators information concerning their own or others performance.</td>
<td>Providing operators feedback has been shown to mitigate appraisal errors and to promote rational AUDs. A combination of multiple forms of performance feedback appears to be more effective than any single type of feedback tested to date.</td>
<td>Beck, Dzindolet, Pierce, Poole &amp; McDowell, 2000; Pomranky, Dzindolet &amp; Peterson, 2001.</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Findings</th>
<th>Pertinent Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attitudes:</strong> Favorable or unfavorable evaluations or reactions to automation.</td>
<td>Surveys typically report substantial variability in respondents’ attitudes toward automation. While some investigators have found that attitudes were reflected in operators’ AUDs, others have shown no association.</td>
<td>Chambers, Jarnecke &amp; Adams, 1976; Goldman, Platt &amp; Kaplan, 1973; Halpin, Johnson &amp; Thornberry, 1973; Lee &amp; Moray, 1992, 1994; McClumpha &amp; James, 1994; Riley, 1994, 1996; Singh, Molloy &amp; Parasuraman, 1993; Tenney, Rogers &amp; Pew, 1998; Weil &amp; Rosen, 1995.</td>
</tr>
<tr>
<td><strong>Trust:</strong> Automation is seen as trustworthy to the extent that it is predictable, dependable, and inspires faith that it will behave as expected in unknown situations.</td>
<td>The operator’s trust in the machine is an important variable determining reliance on automation. However, AUDs can best be predicted by taking both the operator’s trust in the device and their confidence in manual control into account. The comparison of the automated and manual alternatives has been called perceived utility.</td>
<td>Cohen, Parasuraman &amp; Freeman, 1998; Dzindolet, Beck, Pierce &amp; Dawe, 2001; Lee &amp; Moray, 1992, 1994; Moray, Inagaki &amp; Itoh, 2000; Mosier &amp; Skitka, 1996; Muir, 1987, 1994; Singh, Molloy &amp; Parasuraman, 1993; Tan &amp; Lewandowsky, 1996.</td>
</tr>
</tbody>
</table>
useful scheme for organizing the literature, identifying the causes of non-rational AUDs, and suggesting new areas of research. In sharing this model, we are inviting you to try it on, and if the glove fits, use it as a stimulus for your own research.

**Directions for Future Research**

**Recognition of Alternatives.** Dozens of times each day, people perform tasks, never considering whether automation or manual control offers the better way of conducting the activity. Often, automation is assumed to be the superior option without actually comparing its benefits to those of a manual or less technologically complex alternative. Conversely, people frequently cling to manual control, oblivious to the availability of more useful automated devices. When persons do not recognize that there are automated and manual alternatives, they are in effect making AUDs, but without a full appraisal of alternatives. As the AUD Model suggests, some sub-optimal decisions occur simply because operators are not aware that automated and manual alternatives are available.

Although failure to recognize both automated and manual options can produce inefficiency, shortcutting the appraisal process often saves time and energy. To better understand AUDs, we need to determine the conditions that promote abbreviated and more extensive appraisals. Research on decision-making in naturalistic settings (e.g. Beach, Chi, Klein, Smith & Vicente, 1997; Klein, 1997; Salas & Klein, 2001; Zsambok & Klein, 1997) has furnished important information on this issue. If it is assumed that evaluating two options requires greater energy than assessing one, then predictions can also be derived from Shepperd’s Expectancy-Value Theory (1993, 1998; Shepperd & Taylor, 1999). Application of Shepperd’s theory suggests that operators will be most likely to compare automated and manual control when they expect that their efforts are necessary for the group to succeed, believe that a successful performance will lead to a positive outcome, and the value of the outcome outweighs the costs of working hard.

**Appraisal of Options.** As this review indicates, most of the research on AUDs has examined the appraisal process. Still, we are just beginning to discover the variables that bias appraisal and produce misuse and disuse. In this chapter we hypothesized that in addition to automation bias, polarization biases, self-serving biases, illusionary correlations, illusions of control, and the availability heuristic affect AUDs. The individual and social cognitive literatures are rich in other biases that will someday be shown to determine operators’ preferences for automation or manual control. Researchers should not only catalog the many
biases that affect appraisal, they need to identify the conditions in which biases are most likely to occur and develop strategies for minimizing their impact.

Action Selection. The studies by Hawkins, Dzindolet, and Beck (1999), Moes et al. (1999) and Pomranky et al. (2001) demonstrate that non-rational AUDs are not solely due to recognition or appraisal errors. In the feedback conditions, their participants were aware that the automated and manual alternatives were available and knew the utilities of each option. Nevertheless, most operators made action selection errors and refused to let the machine determine their extra credit. If future studies provide a conceptual replication of these findings, we will need to discover the basis of this deep-seated anti-automation bias. One possibility is that an unwillingness to employ obviously helpful automation is sometimes a reaction against change in our society. A hypothesis that deserves examination is that action selection errors are more likely to be observed when the automation threatens the operator’s self-esteem, identity, or occupation.

A Call for Qualitative Investigations. Almost all of the variables known to affect AUDs (see Table 2) were discovered through experimental, survey and other quantitative procedures. While quantitative methodologies will remain central to scientific study, it would behoove us to remember that they essentially yield a picture of what the researcher, rather than the operator, regards as important. Even when operators’ views are assessed, we typically restrict their responses to the topic of our current interest. By asking them about a single or several variables, like trust or perceived utility, we may overlook factors that may have a more significant influence on AUDs.

There is much to gain by complementing our quantitative with more qualitative research approaches. A simple and useful idea that we’ve used in our laboratories (e.g. Dzindolet, Pierce, Beck & Dawe, in press) is to pose an open-ended question. Ask operators why they chose the manual or the automated option. Then pursue that inquiry, exercising as little control as possible over the direction of questioning. The result will be a broader more complete depiction of AUDs, one that reflects the perspectives of the operators.

Need for Interventions to Reduce Misuse and Disuse. Research on AUDs has conclusively shown that misuse and disuse often have a detrimental impact on the performance of human-machine teams. Although most inappropriate AUDs are not life threatening, accident analyses attest to the tragic consequences that misuse and disuse can produce. Fortunately, many sub-optimal AUDs are amenable to training. For instance, pilots lacking visual guideposts are instructed
to use their instruments to avoid disorientation. Beck et al.’s (2000) assessment of the effects of various types of feedback also demonstrated that non-rational AUDs can be controlled. As decision aids come to occupy a more prominent place in the emerging society, researchers will need to develop more effective strategies to prevent recognition, appraisal and action selection errors.

**A Message to System Designers and Operator Trainers**

In contrast to the extensive literature on hardware and software, relatively little information is available on the variables determining automation use. As a result, we are much better at creating automated devices than using them. Fortunately, the research on AUDs and other aspects of automation usage is sending a powerful message to system designers and developers of operator training programs.

New automation often fails to produce expected gains because system designers treat the operator as just another switch or sensor. Although never appropriate, the inadequacies of a mechanical model of operator functioning become increasingly apparent as more intricate tasks are automated. Assuming that the human operator will respond in a mechanical and easily predictable manner is of minor consequence if the goal is to develop an automatic window opener. The actions of operators, whose task is to detect mistakes made by a machine, are less certain; their prediction and control require a more thorough consideration of thought processes and behavior. Still greater understanding of operator performance is needed if persons are to effectively collaborate with decision aids and derive solutions to novel problems.

The inability of intelligent persons to benefit from simple decision aids like the contrast detector emphasizes the need to train operators, not only in the mechanical operation of these devices, but in the wise application of the information that they provide. The research described in this chapter is the beginning of what will be an ever-increasing effort to assess the variables that determine AUDs and produce misuse and disuse. If automation is to enhance productivity and happiness, then the introduction and employment of intelligent machines must be based on a thorough understanding of recognition, appraisal and action selection errors. Improvements in the reliability of decision aids will be a hollow achievement unless our knowledge of hardware and software is matched by an equally sophisticated comprehension of operators’ decision-making processes.

**ACKNOWLEDGMENTS**

The Army Research Laboratory, Human Engineering Directorate, funded much of the research performed in the laboratories of Dzindolet, Pierce and Beck.
The authors would like to thank the following students whose efforts and dedication made this work possible: Theresa Alf, Griffen Angel, Gustavo Araujo, Leslie Barnett, Thomas Blalock, Lee Ann Brown, Kera Cowley, Nicholas Getzen, Greg Greer, Christina Grounds, Jamie Hawkins, Chris Knapp, Kelly Knox, Traxler Littlejohn, Julie Martin, Rebecca McDowell, Marielle Moes, James Nippert, Katherine Piatt, Regina Pomranky, Andrew Poole, Lori Purcell, Jennifer Stansbury, and Alson Wheeler.

REFERENCES


Moes, M., Knox, K., Pierce, L. G., & Beck, H. P. (1999, March). Should I decide or let the machine decide for me? Poster presented at the meeting of the Southeastern Psychological Association, Savannah, GA.


Philip A. Ikomi

ABSTRACT

Aircraft accidents involving airplanes of the three major airplane manufacturers in the database of the National Transportation Safety Board (NTSB) on the World Wide Web were examined for causes. Regression analysis showed that the contribution of automation problems to the total accidents within the 16-year period examined was not statistically significant. However, the contributions of mechanical problems to the overall accident profile was statistically significant $F(1,14) = 5.02$, $p = 0.041844$. Some implications of the findings of the study are discussed.

Without doubt, cockpit automation has been of some benefits to the aviation industry (e.g. Wiener, 1985, 1988). Indeed cockpit automation has given rise to more precise flying, more flexibility in making approaches and executing landings in weather situations that were quite daunting to manual flying and
would have been impossible without automation. In its current state, cockpit automation could enable landing in zero visibility. In fact, automation could help a pilot detect low level wind shear activity at a destination airport and carry out a missed approach thus avoiding an accident. Nonetheless, cockpit automation has also increased workload (Wiener, Chidester, Kanki, Palmer, Curry & Gregorich, 1991), has introduced new roles for the pilot (Sarter & Woods, 1992), and new tasks that the pilot has to learn. In addition, researchers have described several different kinds of errors resulting from the man/machine interaction involved in the use of cockpit automation (e.g. Mosier, Skitka, Heers & Burdick, 1998; Parasuraman, Molloy, & Singh, 1993; Sarter & Wood, 1995).

Mosier et al. (1998) described what they called automation bias which are errors that result from the use of automated cues as short cuts to replace the normal “vigilant information seeking and processing” (p. 48). The errors that are subsumed under automation bias include basically two types of errors, commission errors, and omission errors. An automation commission error according to Mosier et al. occurs when decision makers wrongly take instructions from automated aids. This is usually done even when other indicators show contrary evidence. On the other hand, an automation omission error occurs when decision makers do not take appropriate action because they were not informed by the automated systems in the cockpit. One notable evidence of this kind of error as pointed out by Mosier et al. was the loss of power on engine no. 4 of a China Airlines Boeing 747 SP while in cruise. As a result of the loss, the autopilot made the necessary compensation to correct for the loss of thrust from that engine. Unknown to the pilots is the fact that the airplane had lost an engine. When they disconnected the autopilot for some other reason, the correction being applied for the loss of the no. 4 engine was suddenly lost, and the airplane was suddenly in an unusual attitude with the resultant loss of several hundred feet in altitude and temporary loss of control by the pilots. This error resulted in damage to the airplane.

In their article, Parasuraman, Molloy and Singh (1993) described what they called complacency error which results from an over-reliance on the automated devices within the airplane cockpit for task performance. Complacency errors have also been found to result in accidents.

Finally, Sarter and Woods (1995) have described mode awareness errors. They suggested that the proliferation of modes needed to accomplish performance at different levels of automation has resulted in the need for operators to be aware of the particular mode the automated device is in at any moment in time. This is necessary so that the operator can apply an appropriate action for each mode since an action that is appropriate at one level of automation may not be appropriate at another level of automation. They pointed out for
instance, that there are five ways to change altitude in one of the more automated aircraft in flight. One example of an accident that resulted from mode awareness error, was the Indian Airlines accident in Bangalore, India (viz. Lenovitz, 1990). In that accident, the co-pilot who was the pilot flying was not aware of the consequence of being in the mode selected. The warning from the captain of the flight came too late.

Accidents resulting from mechanical problems with an aircraft come in various forms as well, but usually they are results of failures of one mechanical device or the other. For instance, there have been instances of metal fatigue resulting in the failure of landing gear which have collapsed on landing, engines falling off pylons, and flap drive shafts failing at critical moments. One can think of several other instances of mechanical failures that have resulted in accidents. If one were to use the availability heuristic (Tversky & Khaneman, 1982) in determining the probability of accidents one would probably say that mechanical accidents are more probable than automation related accidents. However, one could also say that automation related accidents are more probable if one has just finished reading about the various types of accidents that could result from the use of automated devices in airplane cockpits. A more appropriate approach would be to actually determine how many automation related accidents have occurred over a period of time and count how many mechanical related accidents have occurred within the same period of time. In this way one can determine the relative likelihood of an automation related accident compared to a mechanical accident.

Although many researchers have documented various automation related errors, some of which have led to accidents that have occurred since the advent of automated devices in airplanes as discussed earlier, the relative frequency of such accidents has not been investigated to this author’s knowledge. I believe that it would be prudent to determine the relative frequency of such accidents in comparison to the more common and well established mechanical accidents. Such a comparison will aid in determining how concerned the aviation community should be about increases in automation related accidents. For instance, if the frequency of automation related accidents far outstrips that of mechanical accidents, one would be very concerned about mitigation of such accidents. However, if the frequency of automation accidents is the same or less than mechanical accidents, the concern would not be so great as it would indicate that the use of automation devices in airplanes is at a level of safety that equals or surpasses the mechanical safety of airplanes.

The study reported here was an attempt to compare the relative frequencies of occurrence of automation related accidents and mechanical related accidents. According to the National Transportation Safety Board (NTSB) (NTSB, 2001)
the accident rate of commercial airliners has remained unchanged for the past two decades. It is possible however, that this general outlook does not describe the individual facets of a multifaceted variable like accident rates. One would agree that there is a multitude of causes to an accident and by the same token a number of variables determine the rate of change of accidents over a number of years. Thus although the overall rate of accidents may be the same over two decades as suggested by the NTSB, individual variable contribution to the accident rate may exhibit variability reflecting the degree of its contribution per year. It is possible to analyze the various accident causes and determine the contribution of each cause to the accidents occurring per year. Through such an analysis, one could determine whether or not these variables have remained unchanged as the overall accident rate has remained the same for the past two decades. The research that is reported here concerned how this kind of analysis was applied to the accident data on the NTSB database from January 1983 to January, 1999, focusing on the impact of automation and mechanical related causes to the total accidents.

Much research money is expended each year by both the Federal Aviation Administration (FAA) and NASA on the issue of automation. Some researchers in this branch of human factors suggest that automation has created more problems than it solves. Furthermore, these researchers fear that if procedures and other avenues of dealing with the problems of automation are not put in place, we could be having very serious accidents in the years ahead. Thus I felt the need to look at the contributions of automation to past accident data. Another point addressed in this study was the extent of the contribution of mechanical problems to accidents in the past two decades. My thoughts were that it is possible that airplanes are being built more reliably and therefore we may be having fewer engine or component failures than in the past. The NTSB accident data could give us answers to those questions.

METHOD

This was an archival study involving the examination of recorded information on airplane accidents kept by the National Transportation Safety Board (NTSB). Since there were a lot of different airplanes involved in flights under various rules and operations, the decision was made to limit the accidents examined to a manageable few. The airplanes that automation researchers seem mainly concerned with are those that carry passengers and cargo and are involved in scheduled, regular departures worldwide. Most of these airplanes are large and are manufactured by three main manufacturers, the Boeing Commercial Airplane Company, a United States based company, the McDonnell Douglas
Airplane Company, also based in the United States, and the European based Airbus Industrie. It was therefore decided to look at airplanes involved in scheduled transportation of passenger and cargo worldwide and to examine airplanes built by the three aforementioned manufacturers involved in those services.

This decision led to the examination of 348 aircraft accident reports involving scheduled passenger gas turbine powered airliners of the Boeing, McDonnell Douglas, and Airbus companies recorded in the database of the NTSB from January, 1983 to January, 1999, obtainable from the World Wide Web (NTSB, 2000). These airliners were involved in flights the FAA classified as Part 121 and Part 129 Scheduled operations. Part 121 scheduled operations are those of the domestic and foreign operations of U.S. registered flag carriers while Part 129 are the scheduled operations of foreign flag carriers.

**Procedure**

In this study, air carrier events classified as accidents by the NTSB, involving scheduled gas turbine powered passenger airliners of the Boeing company, McDonnell Douglas airplane company, and the Airbus company were examined over the period, January, 1983 to January, 1999. The accidents were examined with a view to determining the probable causes based on the NTSB’s identified causes and the interpretation of the reports based on the researcher’s knowledge and experience in aviation.

The NTSB definition of accidents has been adopted in identifying accident aircraft for the study. According to the NTSB, any event that occurs between the time an aircraft is being prepared for a flight and the time it comes to a stop at its final destination that results in a passenger or personnel injury or death or results in any damage to the aircraft is an accident. An accident is to be differentiated from an incident. An incident is an occurrence during the same period in which no injury or death occurs and no damage is sustained by the aircraft. For example, if there was a near miss between two airplanes in the air as reported by air traffic control (ATC) and no one in either airplane sustained injury, then the event will be reported as an incident. However, if there was a near miss and the pilot of one of the aircraft took evasive action which resulted in a flight attendant (FA) being thrown off balance while serving coffee; and the coffee spills on a passenger causing burns, then the aircraft has had an accident. This accident will be reported with such entries as the date of occurrence of the accident, the time of occurrence, the meteorological conditions at the time of occurrence, the airplane registration and nationality marks, the airplane type, and series, the engine manufacturer, and the place of occurrence as well as the number of fatalities, if any, the number seriously
injured and the number of those not seriously injured, among other details. The report also includes a narrative and the probable cause of the accident is stated in summary and contracted form at the end of the narrative. These reports were easy to examine, most of them being only one page long.

In this study aircraft accidents were classified in three categories. Classifications were based on the entire report, not just the probable cause given by the NTSB only. In some of the accident reports, reading the narrative revealed probable automation involvement which was sometimes not given as a probable cause by the NTSB. An accident was classified as “Automation” accident if there was any probable involvement of automation in its occurrence. Automation involvement could take any form, including failure of an automated component (e.g. autopilot), failure of the pilot to properly use an automated component or procedure, or malfunction of an automated component. An accident was classified as “Mechanical” if there was probable involvement of a mechanical failure or malfunction in the accident. An accident could be in both mechanical and automation categories if the accident involved a mechanical failure of an automated component, for instance, failure of the take-off configuration warning device. An automatic warning comes on if an airplane having the device is not properly configured for take-off. The failure of this warning constitutes an automation problem. However, its failure represents a mechanical problem as well. Therefore, it would be classified under both mechanical and automation categories. Any other probable cause of an accident that does not fall under these two categories was classified “Other”. The author classified all the 348 cases into the three categories. Subsequently the reliability of the author’s categorization was assessed by a process of inter-rater reliability assessment. An inter-rater reliability coefficient was calculated based on 46 cases chosen at random from the NTSB accident database and matched with 46 cases classified by the author. Initially 50 cases were chosen, using a table of random numbers, from the sampling domain which included all the relevant accidents in the period 1983 to 1999, but four cases were unusable. A student was briefed on the criteria the author used in classifying the cases into “automation”, “mechanical” and “other” and asked to classify the 46 cases into the same three categories used by the author. The categorization was cross-tabulated and the correlation between the categories given by the two raters calculated. The correlation between the classifications used by the author and the student was, $r = 0.872$ and was significant at the 0.01 level of significance.

The automation related accidents were further classified into 5 categories as detailed opposite:
(1) Procedural – these were problems that resulted from inadequate use of procedures or the improper execution of the relevant standard operating procedures (SOP).

(2) Mechanical – this kind of problem resulted from a mechanical failure or problem in the automated device.

(3) Inadequate knowledge – inadequate knowledge of the device or the functioning of the device as exhibited by the pilot in interaction with the system.

(4) Mode utilization problems – these problems resulted from the operator not being aware of the mode in which the device was functioning at the time of the error.

(5) Design problems – errors of inadequate design. This classification scheme has some rationality to it.

For instance, the procedural classification draws from the procedural school of thought in crew resource management. This school of thought believes that errors can be eliminated or reduced by teaching pilots the procedures to use; and that if pilots use the proper procedures correctly, errors would be reduced. This school of thought is based on research from George Mason University and they have developed the Advanced Crew Resource Management (ACRM) (Seamster, Boehm-Davis, Holt, Schultz & Harman, 1998). The mode utilization category draws from the mode awareness errors of Sarter and Wood (1995).

RESULTS

The total number of reported scheduled air carrier accidents involving the gas turbine passenger airliners of the three major manufacturers within the period, January, 1983 to January, 1999, was 384. Of this number, 36 did not have probable causes assigned either because the reports were incomplete or they were foreign registered airliners. Additionally, reports were not issued if the accidents involved foreign registered airliners outside the jurisdiction of the NTSB in foreign lands. These foreign accidents will not be discussed any further in this report. The variety and number of airplanes involved in the accidents under review are shown in Table 1.

These accidents were responsible for 1149 fatalities within the 16-year period examined. An examination of the data shows that automation accidents were the lowest in number for the 16-year period (see Fig. 1 and Fig. 2). There were 15 automation related accidents, about half of which were also classified as mechanical, that is, a mechanical problem with the automated device contributed to the accident. Seven of the 15 automation related accidents were due to a
### Table 1. Scheduled Airliners Involved in Accidents Under Review.

<table>
<thead>
<tr>
<th>Aircraft Type</th>
<th>#</th>
<th>Aircraft Type</th>
<th>#</th>
<th>Aircraft Type</th>
<th>#</th>
<th>Aircraft Type</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>A300</td>
<td>13</td>
<td>B707</td>
<td>4</td>
<td>B757</td>
<td>25</td>
<td>DC8</td>
<td>8</td>
</tr>
<tr>
<td>A310</td>
<td>1</td>
<td>B727</td>
<td>55</td>
<td>B767</td>
<td>19</td>
<td>DC9</td>
<td>51</td>
</tr>
<tr>
<td>A320</td>
<td>3</td>
<td>B737</td>
<td>73</td>
<td>B777</td>
<td>2</td>
<td>MD11</td>
<td>9</td>
</tr>
<tr>
<td>A340</td>
<td>1</td>
<td>B747</td>
<td>44</td>
<td>DC10</td>
<td>19</td>
<td>MD80</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mechanical problem with automation in the aircraft. Regression analysis showed that the automation contribution to accidents in the 16-year period was not significant. Automation accounted for an adjusted $R^2$ of 0.03, a contribution amounting to only 3% of the variance in the total number of accidents. This value was not a statistically significant contribution to the accident data, $F(1,14) = 1.47, p = 0.245659$.

Fig. 1. Total Accidents in each Category (1983–1999).

Fig. 2. Total Accidents in each Category per Year.
A closer look at the characteristics of the automation accidents showed that they could be partitioned into five sub-categories – those resulting from mechanical failures earlier mentioned, those due to misapplication of procedures by the flight crew, design problems, crew knowledge problems and mode utilization problems. There were five automation accidents resulting from misapplication of procedures, one design problem automation accident, one inadequate crew knowledge problem, and one mode utilization problem. A look at the contribution of mechanical problems to accidents in the last 16 years shows that mechanical problems were responsible for an adjusted $R^2$ of 0.21 or 21% of the variance in the total number of accidents in the period under review. This was a statistically significant contribution, $F(1,14) = 5.02$, $p = 0.041844$.

To determine if the reliability of engines and components in the last 16 years has changed or remained the same, one can assume that the number of mechanical accidents will begin to change noticeably from about the middle of the period. If engines are being built more reliably, the number of mechanical related accidents in the second half of the period will be fewer than the number of mechanical related accidents in the first half of the period. A $t$-test of the mean number of mechanical accidents in the first eight years and the mean number in the last eight years of the period was done: $t(13df) = -0.82$, $p = 0.21$ (one tailed). This test did not show a significant difference between the mean mechanical related accidents in the first half of the period under review and the second half of the period.

The “other” causes category was the largest contributor to the accidents, accounting for an adjusted $R^2$ of 0.92 or 92% of the variance in the total number of accidents occurring in the period. This was the greatest contribution of all three and was statistically significant, $F(1,14) = 154.59$, $p = 0.0000$.

Because the total number of accidents included many considered trivial, I decided to look at accidents that resulted in loss of lives, not in terms of the number of lives lost, but just that lives were lost in the accident concerned. In this realm, automation contributed nothing to reported fatal accidents examined in this study in the last 16 years (see Fig. 3). There were no fatal accidents involving automation problems in the period. Mechanical problems however, accounted for 65.7% of the variance in the fatal accidents within the period while the “other” category contributed 88.3% of the variance in the total fatal accidents reported by the NTSB among the airplanes examined in this study.

The analysis of the contribution of automation problems to accidents suggests that procedural problems were the second most prevalent (33.3% of all automation problems), next only to mechanical automation problems (46.7%).
DISCUSSION

The contribution of automation related problems to the total number of accidents occurring in the 16-year period was statistically non significant. On the other hand, the contribution of mechanical problems to accidents in the period was statistically significant. That procedural problems accounted for most of the automation problems suggests that procedural noncompliance (Helmreich & Merritt, in press) could be a major issue in the operation of automation in airplanes. It is quite surprising that automation did not contribute to any fatal accidents among the transport category airplane accidents reported by the NTSB. (There have been some notorious automation accidents within the period, but unfortunately, the probable causes of the accidents were not reported in the NTSB database, as they were foreign registered aircraft accidents). This lack of contribution by automation related problems to fatal accidents among U.S. registered aircraft should come as a surprise also to those funding automation research. Not only have automation problems not contributed significantly to total accidents, they have not contributed at all to fatal accidents in the last 16 years here in the United States.

Mechanical problems achieved statistical significance. Mechanical problems are found in all areas of operation of flights. Although it is a separate category, it moderates the impact of automation accidents in the overall accident picture. If one were to remove mechanical automation problems from the total automation accidents, automation would account for less than half its present
non-significant contribution. Thus any statement as to the contributions of automation to the accident records in the 16-year period, has as its limitation, the reliability of the component parts of the automation system.

The reliability of engines and components has not changed significantly during the period as suggested by the $t$-test of the first half of the period compared to the second half of the period in review.

This study has not shown whether or not automation has helped in reducing accidents. It is not possible to determine the impact of automation in the reduction of accidents from existing accidents because there might have been accidents that did not take place because of automation and accidents that have occurred because of automation without automation being implicated. However, this study has given us some insight into the dynamics of accidents and their causes. One can suggest that the aviation community does not have to be overly concerned about automation related accidents because as the numbers show, they are far less than mechanical related accidents. The figures thus seem to suggest that safety in the use of automated components seems to parallel or even exceed the safety of mechanical components in a commercial airliner.

Some limitations to the results of this study relate to the classification process used. Although the categories seem pretty clear, different individuals may categorize differently using the same categories. This was clearly demonstrated by the reliability figure obtained (0.872) which, although high, was not perfect.

ACKNOWLEDGMENTS

I want to thank Ore Soluade of Cisco for a valuable discussion of the statistical analysis for the study reported here. I also wish to thank Robert W. Holt of George Mason University for valuable assistance relevant to my choice of an analytical procedure for the determination of the reliability of my categorization. Finally, I would like to thank my undergraduate assistant, Ms. Jennifer Tull-Estwick for selecting and categorizing the accidents used for the reliability assessment. Although these people have helped in various ways, I assume full responsibility for the final outcome of this paper.

REFERENCES


5. AUTOMATION AND COGNITION: MAINTAINING COHERENCE IN THE ELECTRONIC COCKPIT

Kathleen L. Mosier

U.S. aviation authorities are particularly curious about how the Gulf Air pilots handled the A320’s complex computer-operated control and navigation systems. The planes’ software is designed to prevent even the most incompetent pilot from accidentally launching the plane into a fatal stall or dive. But records indicate that some Airbus crashes occurred when pilots misjudged the planes’ limitations or made errors entering data into the A320’s computer system. And because the system is so complicated, U.S. experts say, if something goes wrong only a mathematical genius could figure out the problem (Mark Hosenball, Newsweek, September 4, 2000).

Technological advances in aviation and the increasing sophistication of the automated cockpit have resulted in profound changes in the National Airspace System (NAS) and, more importantly, in the tasks of the human operator within it. For example, the visual task of the pilot has evolved from a focus on perception of aircraft position with respect to terrain, obstacles, clouds, etc., to a monitoring of systems and cockpit displays that give this information. The flight control task, once a “stick and rudder” process, that is, a “systems problem involving the characteristics of the aircraft, the sensing of information necessary for control, the appropriate control laws, and the mechanisms for generating the appropriate control forces and moments” (Baron, 1988), is today essentially an automation problem involving programming and monitoring skills. These changes have been accompanied by the demand for a change in the pilot’s cognitive activity from perception → response to thinking, judging, and deciding.

Advances in Human Performance and Cognitive Engineering Research, Volume 2, pages 93–121. Copyright © 2002 by Elsevier Science Ltd. All rights of reproduction in any form reserved. ISBN: 0-7623-0864-8
Unfortunately, current understanding of the modern flying task in the context of the automated cockpit is limited, because the evolution in aviation technology has not been accompanied by a parallel evolution of cognitive theoretical models and frameworks within which to study it. Rather, investigators have utilized a piecemeal approach, and have parsed out facets of the task – such as situation awareness, risk assessment, automation bias, mode confusion – and studied them in order to identify and prescribe remedies for associated errors. Because aviation researchers have employed a fragmented approach and have gained an incomplete understanding of the cognitive processes involved, we are unable to make more than fragmented predictions.

Modern researchers have acknowledged and challenged the lack of a unified approach. Parasuraman (2000), for example, cites several qualitative and quantitative models of human-automation interaction and notes that “An important future research need is the integration of qualitative and quantitative models . . .” Maurino (2000) laments that “. . . aviation research on human judgement and decision-making has until recently been conducted out of context and it has ignored the natural component of the decision-making process . . . Likewise, errors have been considered to emerge from either the technology or the human, but seldom from the joint human-technology system . . .” In our approach, then, we have not progressed far since Wickens and Flach’s (1988) analysis of the information processing approach to human cognition in aviation: “. . . we have missed the big picture” (p. 148).1

The implications of changes in aviation – particularly advances in automated systems and displays – in terms of the pilot’s task are enormous, and have been discussed at length from many perspectives (see, e.g. Billings, 1996; Rasmussen, Duncan & Leplat, 1987; Wiener, 1989; Parasuraman & Mouloua, 1996; Parasuraman, Molloy & Singh, 1993; Norman & Orlady, 1989). As technology within the cockpit became increasingly complex, it became clear that the information-processing approach that had been accepted and utilized for several decades had “basic and serious shortcomings” (Hollnagel, 2000, p. 66). The shift from active control to monitoring has changed not only the type of cognitive activity demanded, but also the associated consequences – and this change demands a new theoretical perspective. Most importantly, in terms of theoretical implications, the automated cockpit brings cues that were in the outside environment into the cockpit, and displays them as highly reliable and accurate information rather than probabilistic cues. This changes the goal of pilot cognition from correspondence, or empirical accuracy in using probabilistic cues for diagnosis, judgment, and prediction, to coherence, or rationality and consistency in diagnostic and judgment processes.
What is needed in aviation research, then, is a theoretical framework that defines essential components of the pilot’s task, and recognizes that: (a) flying in today’s automated aircraft is essentially a cognitive task; (b) the goals of cognition, as well as the cognitive tactics and strategies utilized to achieve them, are impacted by features of the operational environment; and (c) it is essential to examine pilot behavior and errors in the automated cockpit in terms of a match or mismatch between the cognitive tactics elicited by the environment and adopted by pilots, and the cognitive strategy required by the task. The premise of this chapter is that such a framework can be found in correspondence and coherence, complementary metatheories of judgment and decision making, in combination with the Cognitive Continuum Theory of judgment (CCT; e.g. Hammond 1996, 2000; Hammond, Hamm, Grassia & Pearson, 1997). The purpose of this chapter is to describe changes that have occurred during the evolution of aircraft automation and the implications of these changes for pilot cognition, and to introduce these theoretical concepts as a potential unifying framework within which to examine pilot cognition in aviation.

THE EVOLUTION OF THE PILOT’S TASK: FROM (VISUAL AND KINESTHETIC) SENSE TO (COGNITIVE) SENSIBILITY

Piloting an aircraft used to be a very “physical” task. First generation aircraft were highly unstable, and demanded constant physical control inputs (Billings, 1996). Perception → response was the key process involved in piloting them. Early aviation research focused on troubleshooting manual control and operational problems. In early days of aviation, the goal of the pilot, correspondence competence (accuracy), was achieved through the senses via visual and kinesthetic perception – what has been referred to as “contact” flying (Hopkins, 1982). For pilots, the emphasis was on accurate judgment of objects in the environment – height of obstacles in terrain, distance from ground, severity of storm activity in and around clouds, location of landmarks – and accurate response to them (e.g. using the controls to maneuver around obstacles or storm clouds, or to make precise landings).

Pilots avoided situations that would put the accuracy of their senses in jeopardy, such as clouds or unbroken darkness. For example, early airmail pilots often relied on railroad tracks to guide their way, and mail planes had one of the landing lights slanted downward to make it easier to follow the railroad at night. Later, a system of beacons and gas lights created a 902-mile illuminated
airway for nighttime flight (Orlady & Orlady, 1999). As pilots gained experience, more accurate perception resulted in more accurate response. The focus that dominated aviation research for many years, perception and human information processing, has been consistent with a this era of aviation history. Limitations of the human as perceiver, such as spatial disorientation or the degradation of vision at night (e.g. Leibowitz, 1988), or as information processor, such as attention and memory limitations (e.g. Wickens & Flach, 1988), have been key topics of research. Much of the research was geared toward overcoming these limitations, often by bringing information into the cockpit in formats that took into account human deficits and compensated for them.

As aircraft evolved, flying became physically easier as parts of control task were automated (e.g. through use of an autopilot), and automated systems began to perform many of the flight tasks previously accomplished by the pilot. The demand for all-weather flight capabilities resulted in the development of instruments that would supposedly compensate for any conditions that threatened to erode pilots’ perceptual accuracy. Conflicts between visual and vestibular cues when the ground was not visible, for example, could lead to spatial disorientation and erroneous control inputs – but an artificial horizon, or attitude indicator, within the cockpit, could resolve these conflicts. Additionally, more and more information inside the aircraft supplemented or replaced cues outside the aircraft (e.g. altitude indicator, airspeed indicator, alert and warning systems). Many of the cues pilots had traditionally used to maintain an accurate perception of and position in the world (i.e. flight control and navigational accuracy) were altered or removed. For example, instruments that became increasingly reliable allowed aircraft able to operate in low visibility conditions, eliminating out-the-window visual perception cues. In some current aircraft, “fly-by-wire” controls, which provide little or no tactile feedback on thrust setting, have replaced conventional hydraulically actuated control columns (see Billings, 1996, for a complete description of the evolution of aircraft automation).

In modern, high-tech aircraft, the flying task is much more cognitively than physically demanding. It is much more a matter of cognitive sensibility than of sensing. In contrast to earlier aviators, glass cockpit pilots can spend relatively little of their time looking out the window or manipulating flight controls, and most to all of it focused on integrating, updating, and utilizing information inside the cockpit. This represents a profound change in both the pilots’ operational environment and the attendant cognitive requirements to function successfully within it. Specifically, the change in task must be accompanied by a change in pilot tactics and strategic goals – from intuitive correspondence to analytical coherence.
STRATEGIC GOALS OF COGNITION IN THE OPERATIONAL ENVIRONMENT: CORRESPONDENCE/COHERENCE

Correspondence

The goal of correspondence in cognition is empirical, objective accuracy in human judgment. Correspondence competence refers to an individual’s ability to accurately perceive and respond to multiple fallible indicators in the environment (e.g., Brunswik, 1956). Correspondence is a relatively “natural,” adaptive process; “. . . we exercise our correspondence judgments almost instantaneously without thinking, just as we correctly perceive the world around us without thinking, without making strong, conscious demands on memory . . .” (Hammond, 2000, p. 35). A pilot, for example, exercises correspondence competence when using cues outside the cockpit to figure out where he or she is, or judging height and distance from an obstacle or a runway. A weather forecaster performs a correspondence judgment when predicting tomorrow’s weather from today’s weather patterns. The measure of competence is the accuracy of the judgment – how well it corresponds to fact. Features of the environment and of the cues utilized will impact the accuracy of correspondence judgments. For example, cues that are concrete and/or can be perceived clearly will facilitate accurate judgments. A pilot will have a relatively easy time judging a 5-mile reporting point when it is marked by a distinctive building. Cues that are murkier, either because they are not as concrete in nature or because they are obscured by factors in the environment, will hinder accurate judgments. The same pilot will have a much harder time judging the report point at night, or when the building is hidden by fog or clouds. Correspondence judgments cannot be made without reference to the “real world,” and are evaluated according to how well they represent, predict, or explain objective reality.

Signal detection theory. Signal detection theory (SDT; e.g., Green & Swets, 1974; Tanner & Swets, 1954), a widely-used correspondence theory, focuses on the presence or absence of a stimulus or event, and the correct/incorrect detection of its occurrence/non-occurrence. Signals and/or noise can be enhanced or degraded to impact probability of detection or error. Cognition and action are evaluated according to the accuracy of detection. In terms of this theory, a pilot looking out the window for other aircraft has four possible outcome options: a hit, or correct detection of an aircraft; a miss, or failure to detect an aircraft that is present; a false alarm, or false perception of an aircraft; or a correct rejection, or the correct perception that no other aircraft are present (e.g., Wickens & Flach, 1988). His or her judgments would be evaluated
according to these outcome options, which are measures of accuracy with respect to the real-world presence or absence of other aircraft.

**Brunswik’s lens model.** A classic description of the correspondence process is contained in Brunswik’s lens model (e.g. 1943, 1956). According to this model, judgments of people and objects are based on “probabilistic cues,” or attributes that can be perceived by an individual and are used to evaluate elements in the environment (e.g. Hammond, 2000). The degree of uncertainty between cues and their criterion is referred to as “ecological validity.” The model has been applied in a variety of contexts, including interpersonal perception and interpersonal learning, policy capturing, weather forecasting, and conflict resolution (see Cooksey, 1996, for a review). The notion of uncertainty was incorporated early on into Wickens and Flach’s (1988) model of pilot decision making. They proposed that the pilot must deal with multiple probabilistic cues, such as visual landmarks, smells (e.g. of smoke), aural messages, more-or-less reliable instruments, etc., in assessing the state of the world. “The cues used for situation assessment may be unreliable (e.g. a weather forecast predicts a 20% chance of thunderstorms), and the projected consequences of an action into the future are uncertain” (p. 127). One source of pilot correspondence error, or inaccurate empirical judgments, is the tendency to respond inappropriately to cues – that is, pilots tend to treat all cues as though they have equivalent reliability and validity, or they utilize cues according to their salience rather than their reliability or validity. “Good” pilot decision makers learn, however, to use these probabilistic cues effectively to make accurate assessments and predictions.

The emphasis of correspondence theories, then, is on the objective correctness of human judgment and the factors that influence it. The decision maker makes a judgment or prediction based on cues or indicators in the environment, all of which are fallible to some degree. The ultimate test of the process is the empirical accuracy of the resultant judgment.

**Coherence**

Coherence theories, on the other hand, focus on the **rationality** of the decision-making process. **Coherence competence** refers to an individual’s ability to maintain logical consistency in judgments and decisions. A pilot, for example, exercises coherence competence when he or she scans the information displayed inside the cockpit to ensure that system parameters, flight modes, and navigational displays are consistent with each other and with what should be present in a given situation. What the pilot strives for is a rationally “good” picture – engine and other system parameters should be in sync with flight mode and
navigational status – and decisions that are consistent with what is displayed. What is important is the logical consistency, or coherence, of the process and resultant judgment. Coherence competence must be acquired artificially – it entails training on what constitutes coherence, and time to think when applying what has been learned – and is more susceptible to time pressure, distraction, and stress (Hammond, 2000). In contrast to correspondence competence, coherence competence is not evaluated by empirical accuracy relative to the real world; the quality of the cognitive process utilized is the sole evaluative criterion.

**Heuristics and biases.** Much of the research on coherence in judgment and decision making has focused on the difficulty humans have maintaining coherence. The “heuristics and biases” research is perhaps the best-known example of a coherence approach to judgment research (e.g. Tversky & Kahneman, 1974; Kahneman, Slovic & Tversky, 1982). Researchers in this tradition compare human judgment, which they have found to be characterized by various heuristics (short-cuts) that individuals use to speed up the decision-making process, against normative or mathematical models. They have found human judgment to be, at best, an approximate process, and, at worst, irrational and subject to systematic biases. The focus of this research is on the (in)coherence of the process, rather than on judgment accuracy. The key issue is not whether heuristics may result in accurate judgments, but rather the notion that they exemplify the flawed nature of the human judgment process.

The use of heuristics is fostered by the tendency of humans to be “cognitive misers,” that is, to take the road of least cognitive effort (e.g. Fiske & Taylor, 1994). Judgments based on heuristics are not rational or coherent, because they are based on incomplete and/or incorrect information. In aviation, the use of heuristics such as availability, the judgment of the likelihood of events dependent on the ease of their recall, or representativeness, the tendency to use the similarity of a past event to a present situation to judge the likelihood that the current event will have the same outcome, has been identified as a factor in incidents and accidents (e.g. Nagel, 1988; Wickens & Flach, 1988).

**Automation bias.** Recently, we have identified automation bias, a flawed decision process characterized by the use of automated information as a heuristic replacement for vigilant information seeking and processing, as a factor in pilot decision errors. Two classes of automation-related coherence errors commonly emerge in highly automated decision environments: (1) omission errors, defined as failures to respond to system irregularities or events when automated devices fail to detect or indicate them; and (2) commission errors, which occur when decision makers incorrectly follow an automated directive or recommendation, without verifying it against other available information, or in spite of
contradictions from other sources of information (e.g. Mosier, Skitka, Dunbar & McDonnell, 2001; Mosier, Skitka, Heers & Burdick, 1998; Skitka, Mosier & Burdick, 1999).

Studies of this phenomenon involved professional pilots flying event scenarios in a high-fidelity “part-task” simulator, that is, a computer-based flight simulator that incorporated the flight instrumentation and dynamics of a multi-engine glass-cockpit aircraft (Mosier et al., 1998, 2001), as well as student participants performing a low-fidelity flight analog task (Workload/PerformANcE Simulation, or Window/Panes, NASA Ames Research Center, 1989; Skitka et al., 1999; Skitka, Mosier, Burdick & Rosenblatt, 2000). In all scenarios, correct information was available to cross-check and detect automation anomalies; that is, to detect disruptions to coherence and correct them. Results of these studies documented comparable automation error rates across sample populations – for both pilots and students, errors rates approximated 55%. These findings suggest that something about the automated environment may be encouraging non-rational cognitive processing and hindering the maintenance of coherence.

This suggestion is further supported by a student study comparing error rates in a low-fidelity flight task with and without a highly, but not perfectly reliable computer (AMA, or Automated Monitoring Aid) that monitored system states and made decision recommendations. When the AMA worked properly, people in the automated condition made fewer errors than those in the non-automated condition. However, the presence of the AMA fostered a tendency to rely on it to detect disruptions and maintain coherence – when it did not work properly, people with an AMA were much more likely than those without the aid to miss specific disruption events, if the AMA failed to notify them of the event (41% vs 3% error rate), and were highly prone to make AMA-prompted commission errors (65% error rate; Skitka et al., 1999).

Student as well as pilot participants in other studies were found to be sensitive to the differential importance of various subtasks, and were more likely to avoid automation bias on those that were more important. Pilots’ errors, for example, roughly reflected the criticality of various flight tasks – they made fewer errors on altitude-related events, for example, than on events involving the communications frequency (Mosier et al., 1998, 2001). An interesting facet of pilot commission errors during a false engine fire event was a phenomenon we dubbed “phantom memory.” Pilots tended to erroneously “remember” the presence of expected cues confirming the presence of an engine fire, thus supporting their subsequent decision to shut down the supposedly affected engine. In terms of coherence, these pilots were misled by their expectations, and perceived a coherent diagnostic representation of an engine fire where only a disjointed picture existed.
In the automated cockpit, it should be noted, more than one person is available to perform tasks and monitor systems. When two people are monitoring system events, it would seem to double the chances that their decision processes would be coherent and rational – that they would detect an anomaly, even if it were not detected by an automated decision aid, or recognize an inconsistent or inappropriate automated recommendation. Many cockpit procedures, in fact, are designed on the premise that crew members will cross check system indicators as well as each others’ actions. What, then, would be the impact of a second crew member in terms of automation bias and automation-related errors? As reported in Mosier et al. (2001) and Skitka et al. (2000), the presence of a second crew member was not sufficient to eradicate automation bias and associated errors. Error rates were not significantly different between one-person and two-person crews. Again, consistent with the previous study, pilot crew errors reflected flight task criticality, and the “phantom memory” phenomenon appeared in flight crews that decided to shut down the supposedly affected engine during the false engine fire event. Interestingly, both of the crews that left the engine running and available for use knew that no other indicators were present, and that the engine fire “story” was not coherent.

Heuristic reliance on automated information was not evident when presented in a format that made other information equally salient. In two separate studies, we examined student and pilot responses to paper-and-pencil decision-making scenarios, and found no systematic preference for automated information (Skitka, 1999; Mosier, Keyes & Bernhard, 2000). When information from non-automated sources was presented with equal salience to automated information, people made decisions more “rationally,” and automation bias effects did not emerge. Rather, participants acted in accordance with perceived risk, and acted to minimize the potential hazard associated with their choices.

Several aspects of the automation bias findings should be noted:

- Results of these studies suggest that something about the automated environment makes operators susceptible to the faulty judgment processes that result in coherence errors. They do not exhibit these errors when the same information is presented in a non-automated environment.
- This susceptibility persists even when two people are sharing responsibility for monitoring, judgment, and decision making.
- The “phantom memory” phenomenon illustrates “the strength of our intuitive predilection for the perception of coherence, even when it is not quite there” (Hammond, 2000, p. 106), and highlights the fact that pilots may not be aware of non-coherent states even when the evidence is in front of them.
Experience with automated systems does not decrease susceptibility to automation bias, as evidenced by equal error rates across student and professional pilot samples. In fact, within the pilot sample, experience was inversely related to the coherence of processes – the higher the amount of experience pilots had, the more automation-related errors they made (Mosier et al., 1998). Automation bias, then, describes a phenomenon that erodes the coherence of pilot cognitive processes. This and other research in heuristics and biases points toward the importance of coherence competence in aviation cognition, and the potential negative impact of lack of coherence. The difficulty inherent in maintaining coherence in the cockpit, however, is also evident from this research. Even the most experienced glass-cockpit pilots exhibit the tendency to rely on automated information without regard to the congruence of other sources of information – in fact, they often do not bother to check all sources. The high internal reliability of automated systems may actually exacerbate this tendency, by making it possible for pilots to perform effectively most of the time despite the use of non-coherent cognitive processes.

The Relationship between Coherence and Correspondence

Three aspects of correspondence and coherence are critical in understanding their role in aviation cognition. First, correspondence and coherence are either/or processes. An individual may alternate between coherence and correspondence, but cannot do both at once (Hammond, 2000). While landing an aircraft, a pilot will switch back and forth rapidly from correspondence to coherence – checking cues outside of the window, glancing inside at cockpit instruments, back out the window – or, in some cases, one crew member will be responsible for coherence and the other for correspondence. A standard landing routine, for example, calls for one pilot to keep his or head “out the window” while the other monitors instruments and makes altitude callouts.

Second, in aviation, correspondence and coherence are complementary processes. Both correspondence and coherence are critical goals. Cognition in aviation demands coherence competence – in the cockpit, in the air traffic control center, in on-ground operations – but cognition in aviation also requires correspondence. Pilots must be able to trust the empirical accuracy of the data used to achieve correspondence; that is, the achievement of coherence must also accomplish correspondence. The pilot may not be able to verify this because he or she does not always have access to either correspondence cues or to “objective reality.” When programming a flight plan or landing in poor weather, for example, the pilot must be able to assume that the aircraft will fly what is
programmed and that the instruments are accurately reflecting altitude and course. When monitoring system functioning, the pilot must be confident that the parameters displayed on the instrument panel are accurate.

Additionally, the cognitive tactics by which pilots may accomplish coherence and correspondence vary on a continuum from intuition to analysis. As described below, the appropriate strategy varies as a function of task and display properties. Perhaps the most critical factor impacting pilot ability to achieve and maintain coherence in the automated cockpit is the degree of match or mismatch between the cognitive tactics elicited by task and display features, and what is required for coherence.

**COGNITIVE TACTICS ELICITED BY THE OPERATIONAL ENVIRONMENT:**

**INTUITION → ANALYSIS**

Coherence and correspondence are the strategic goals of judgment and decision making; *intuition* to *analysis* are cognitive tactics used to achieve them. *Analysis* refers to a “step-by-step, conscious, logically defensible process,” whereas *intuition* typically describes “the opposite – a cognitive process that somehow produces an answer, solution, or idea without the use of a conscious, logically defensible, step-by-step process” (Hammond, 1996, p. 60). Hammond (in preparation) refers to intuition and analysis as the “easy” and “hard” ways to achieve coherence and correspondence.

The notion of a cognitive continuum from intuition → analysis has been thoroughly developed by Hammond (e.g. 1993, 1996, 2000). According to his cognitive continuum theory, intuition and analysis represent the endpoints on a continuum of cognitive activity. Judgments vary in the extent to which they are based on intuitive or analytical processes, or some combination of both. At the approximate center, for example, is *quasi rationality*, sometimes referred to as common sense, which involves components of intuition as well as analysis. During the judgment process, individuals may move along this continuum, oscillating between intuition and analysis – or stopping at points on the continuum. They may shift between intuitive and analytical strategies many times within the same problem (Hamm, 1988). Pilots, for example, may use intuition when gauging weather from clouds ahead, switch to analysis to read and interpret printed weather data, and utilize some combination of the two to judge the safest path, or decide whether to continue on or turn back.

Processes described by any point on the continuum may be used to achieve correspondence or coherence (see Fig. 1). In the aviation context, novice pilots may *analytically* strive for correspondence – accuracy – by using a combination
of cues, rules and computations to figure out when to start a descent for landing. Pilots also learn to use intuitive, pattern-matching processes to assess cues and judge situations. As they gain more experience, the correspondence process becomes more recognitional, and their intuitive assessment of whether the situation “looks right” to start down becomes increasingly effective. In the naturalistic environment, a pilot’s correspondence competence – that is, the ability to utilize probabilistic cues in the environment to assess situations and predict outcomes – increases with expertise. Expert pilots are able to quickly recognize a situation, and may be able to use intuitive processes under conditions that would demand analysis of a novice.

Intuition and analysis differ on several cognitive properties. Intuition is characterized by rapid data processing, with low cognitive control and little to no conscious awareness of processes. It is imprecise but robust, in that accuracy can be achieved despite the inappropriate use of a weighted average organizing principle, and errors tend to be normally distributed. People who make judgments intuitively typically exhibit high confidence in their answer, although they may not be as confident in the method by which it was derived. In contrast, analysis demands high cognitive control and conscious awareness. It requires task-specific organizing principles, and slow processing of data. Analysis is precise but fragile, in that errors occur less often but are likely to be large when they do occur. People who make judgments analytically typically exhibit high confidence in their method, but may not be as confident about their answer (see Hammond, McClland & Mumpower, 1980; Hammond et al., 1997).

Specific task properties are conducive to, or induce, specific modes of cognition, as shown in Table 1. Task environments that contain a large number of cues displayed simultaneously and briefly, high redundancy among cues, perceptual measurement, and a low ability to attain certainty in the task will induce intuitive cognition. In contrast, task environments that contain a small number of cues displayed sequentially and at length, low redundancy among
cues, and the ability to attain certainty in the task will be conducive to analytical
cognition. The precise location of cognitive activity along the continuum, that
is, the amount of intuition and/or analysis used to complete the task, will depend
on which task properties are present, and to what extent (number and amount; 
Hammond et al., 1997).

In a study of expert highway engineers, for example, Hammond et al. (1997),
found that both surface and depth characteristics of a task (i.e. display format
and covert relationships among variables within the task) tended to induce
corresponding cognitive responses. Judgments of highway aesthetics from film
strips representing segments of the highways tended to be performed intuitively.
Judgments of highway capacity via mathematical formulas were accomplished
analytically. Quasi-rational cognition was induced by requiring judgments of
highway safety from bar graphs (Hammond et al., 1997). Perhaps the most
compelling finding of this study was that no one mode of cognition was best

Table 1. Inducement of Intuition and Analysis by Task Conditions.*

<table>
<thead>
<tr>
<th>Task Characteristic</th>
<th>Intuition-Inducing State of Task Characteristic</th>
<th>Analysis-Inducing State of Task Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number of cues</td>
<td>Large (&gt; 5)</td>
<td>Small</td>
</tr>
<tr>
<td>2. Measurement of cues</td>
<td>Perceptual measurement</td>
<td>Objective reliable measurement</td>
</tr>
<tr>
<td>3. Distribution of cue values</td>
<td>Continuous highly variable distribution</td>
<td>Unknown distribution; cues are dichotomous; values are discrete</td>
</tr>
<tr>
<td>4. Redundancy among cues</td>
<td>High redundancy</td>
<td>Low redundancy</td>
</tr>
<tr>
<td>5. Decomposition of task</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>6. Degree of certainty in task</td>
<td>Low certainty</td>
<td>High certainty</td>
</tr>
<tr>
<td>7. Relation between cues and criterion</td>
<td>Linear</td>
<td>Nonlinear</td>
</tr>
<tr>
<td>8. Weighting of cues in environmental model</td>
<td>Equal</td>
<td>Unequal</td>
</tr>
<tr>
<td>9. Availability of organizing principle</td>
<td>Unavailable</td>
<td>Available</td>
</tr>
<tr>
<td>10. Display of cues</td>
<td>Simultaneous</td>
<td>Sequential</td>
</tr>
<tr>
<td>11. Time period</td>
<td>Brief</td>
<td>Long</td>
</tr>
</tbody>
</table>

* From Hammond (1997).
for every task – performance was best when correspondence between task properties and most effective cognitive process was highest.

People will typically prefer intuition over analysis because it is patently easier and quicker, and involves less cognitive effort (e.g. Fiske & Taylor, 1994). Humans “avoid reasoning their way to solutions, and prefer to pattern match” (Moray, 1994, p. 74). They resist the need to calculate or optimize (Rouse, 1983). With respect to modern aircraft, pilots may tend to take short-cuts when possible, relying “. . . uncritically on automation without recognizing its limitations or fail[ing] to monitor the automation’s behavior” (Parasuraman & Riley, 1997, pp. 238–239). In a comparison of pilots and students on an automated task, for example, Riley (1994) found that the pilots were more likely than students to continue to rely on the automation after it had failed. Parasuraman and Riley (1997) have discussed this tendency toward overreliance as “automation misuse.” Pilot trust in electronic systems, a byproduct of their high internal reliability, exacerbates the tendency to delegate analysis as well as task performance to the systems (e.g. Lewandowsky, Mundy & Tan, 2000; Mosier & Skitka, 1996; Mosier et al., 1998). Automation philosophies of some airlines may lead pilots to rely on automated systems to maintain coherence within the cockpit. In fact, pilots sometimes ignore indications of an apparent lack of coherence, thinking that they must be ‘missing something’ that would make the picture OK. These issues highlight the need to tailor the display of information to the desired cognitive response. It is critical to design systems that will aid human metacognition – that is, help pilots to recognize when an intuitive response is inappropriate, and analytical cognition is needed.

**Expertise and cognition.** The knowledge and cognitive abilities that define domain expertise vary according to whether coherence competence or correspondence competence is required. When correspondence competence in the decision making process is required, expertise entails knowing which of the multiple fallible indicators to look for and rely on, and how to use and act upon these indicators. Expertise in a coherence-based task requires knowledge of how a system works, and the ability to describe the functional relations among the variables in a system (Hammond, 2000). Moreover, expertise within a domain does not ensure expertise in *decision making* within that domain. “Cognitive competence takes two forms; first and foremost is *subject matter* competence (often called *domain* competence); the second form is *judgment and decision making* competence. . . . This is an important distinction to grasp, for it is often mistakenly assumed that domain competence assures process competence, or competence in judgment” (Hammond, 2000, p. 32).

Much of the applied work that has been done on expert processes in aviation has focused on correspondence competence. Klein’s model of expert
Recognition-Primed Decision Making (e.g. Klein, 1993; Klein, Calderwood & Clinton-Cirocco, 1986), for example, describes expertise as the ability to identify critical cues in the environment, to recognize patterns of cues, and to understand the structural relationships among cues. Experts are adept at assessing cue validity within specific contexts, and their assessments of a given situation are likely to be more accurate. They are also better than inexperienced decision makers at predicting the outcome of a given decision or action. Chess experts, for example, not only calculate the best position, but also are much better than novices at predicting their opponent’s response. Expert firefighters use cues such as the sponginess of the floor to predict the path of a fire.

Expert pilots typically exhibit high correspondence competence. Their expectations have been shaped by a wide array of experiences, and they utilize these experiences to assess patterns of cues. For example, the pilot may check to see if the view out the window “looks right” with respect to patterns of surrounding and runway lights, and whether the cues match what he or she has previously encountered at this location or point in the flight. Or, the pilot may scan the sky ahead, gauging a safe distance from clouds and picking out a safe route through them. Experts look for familiar patterns of relevant cues, signaling situations that they have dealt with in the past, and base their responses on what they know “works” (e.g. Klein, 1993; Klein, Calderwood & Clinton-Cirocco, 1986). This expertise makes their correspondence judgments and subsequent decisions more likely to be accurate. Training for correspondence competence, in fact, often involves repeated exposure to specific types of situations, via paper-and-pencil exercises or computer simulations (e.g. Klein, 2000; Wiggins, O’Hare & Lods, 2000).

A large body of literature exists extolling the virtues of expertise and expert “intuition” (e.g. Dreyfus, 1997; Klein, 1993). The “intuition” discussed in this literature actually describes the situation-recognition-based decision processes typically utilized by experts. Expert pilots, for example, may use a combination of situation assessment (which is accomplished via pattern recognition), and knowledge gained through experience and past analyses to make judgments that appear to be “intuitive” because they are relatively fast and do not involve computations. Their responses, however, are rooted in a breadth of knowledge garnered from years of analyzing situations and patterns in the environment. Experience, then, is a key factor in improving correspondence competence to the point of “expertise.”

Experience, however, does not necessarily result in greater coherence competence (Hammond, 1996). It may even, in aviation, be counterproductive if accompanied by the tendency to rely on electronic data and systems in an intuitive manner. Riley (1994), for example, found that experienced pilots were
more likely than students to inappropriately rely on automated systems after they (the systems) had failed. As mentioned earlier, more experienced pilots exhibited a greater tendency toward automation bias (Mosier et al., 1998). Experience can also work against a pilot, in that it may induce a “false coherence,” or the tendency to see what he/she expects to see rather than what is there, as illustrated by the “phantom memory” phenomenon found in our part-task simulation data (Mosier et al., 1998, 2001).

In sum, the characteristics of expertise vary according to the cognitive demands that must be met. Correspondence-based expertise, which focuses on using past experience to deal with fallible indicators in a present uncertainty, is not sufficient preparation for the cognitive demands of the automated cockpit. What is required in the electronic, deterministic world of the glass cockpit is analytical coherence.

EVOLUTION OF COGNITIVE DEMANDS IN THE OPERATIONAL ENVIRONMENT: FROM INTUITIVE CORRESPONDENCE TO ANALYTICAL COHERENCE

The task of flying in a high-tech aircraft, then, has evolved from a largely correspondence-based, physically demanding task to a primarily coherence-based, complex, cognitively demanding mental task. This evolution demands an accompanying change in pilot cognition – from intuitive correspondence to analytical coherence.

The early cockpit domain was correspondence-driven in that the achievement of empirical accuracy with respect to the objective environment was a primary goal. The flying task involved integrating probabilistic cues (multiple fallible indicators; see, e.g. Brunswik, 1956; Hammond, 1996) to formulate judgments of position, the feasibility of particular course of action, etc. The kind of cognitive processes utilized were largely intuitive ones – imprecise but robust, enabling rapid data processing and quick judgments, with relatively low awareness of the nature of one’s “reasoning,” but with high confidence in one’s response. Correspondence, however, is no longer the salient cognitive goal for the pilot, because aircraft systems and sensors accomplish most of the “correspondence” tasks. For example, exact distance from the ground can be read from the altimeter – some aircraft even provide auditory altitude callouts during landings. The location and severity of storm activity in and around clouds is displayed on color radar. Today’s aircraft can fly from A to B in low (or no) visibility conditions – once initial coordinates are programmed into the flight computer, navigation can be accomplished without any external reference
cues. Flight management automation allows pilots to program tasks to be performed by automated systems.

The modern flying task is essentially coherence- rather than correspondence-based. The data that pilots utilize to fly can, in most cases, be found on cockpit display panels and CRTs, and are qualitatively different from the cues used in correspondence judgments. They are data, rather than cues – that is, they are precise, reliable indicators of whatever they are designed to represent. The “ecological validity” of the glass cockpit, or the “probabilistic mapping between the environment and the medium of perception and cue utilization” (Flach & Bennett, 1996, p. 74), approaches 1.0. Probabilism has essentially been engineered out of the cockpit through high system reliability.

The cockpit is now an electronic, deterministic environment, in which primary task of pilot is to supervise and monitor systems and information displays to ensure consistency, or coherence, of the “world” and to restore it when disruptions occur. To a great extent, interactions with humans and with the environment have been supplanted by interactions with automated systems:

The development and introduction of modern automation technology has led to new cognitive demands. . . . The result is new knowledge requirements (e.g. understanding the functional structure of the system), new communication tasks (e.g. knowing how to instruct the automation to carry out a particular task), new data management tasks (e.g. knowing when to look for, and where to find, relevant information in the system’s data architecture), and new attentional demands (e.g. tracking the status and behavior of the automation as well as the controlled process) (Amalberti & Sarter, 2000, p. 4).

The importance of handling skills has been replaced by an expectation of management skills, requiring rule-based and knowledge-based cognitive control. The more advanced the aircraft, the higher the demand for coherence competence, or the ability to rationally achieve and maintain consistency among indicators. Accomplishing this demands analytical cognitive processing.

Within the electronic cockpit, pilots are required to demonstrate coherence competence; that is, to think logically and analytically about data and information in their work environment. This involves knowing what data are relevant, integrating ALL relevant data to come up with a “story” of the situation, and ensuring that the story that the data present is rational and appropriate. It also entails recognizing inconsistencies in data that signal lack of coherence, as well as understanding the limits of coherence-based systems and recognizing their strengths and inadequacies. In some cases, for example, coherence systems will warn you of violations – being too low on glide slope – but in other cases, they won’t – going too fast under 10,000 ft.

Many if not most pilot errors in electronic cockpits are failures to detect a disruption in coherence – that is, something in the electronic “story” that is not
consistent with the rest of the picture. Parasuraman and Riley (1997), describing the need for better feedback about the automation’s state, cite “controlled flight into terrain” accidents in which crews failed to notice that they had selected the wrong guidance mode, in part because system indications appeared similar to what they would be when tracking the guide slope (Corwin, Funk, Levitan & Bloomfield, 1993). “Automation surprises,” or situations in which crews are surprised by actions taken by automated systems (Sarter, Woods & Billings, 1997), are other instances of coherence disruption. These occur when pilots have an inaccurate judgment of coherence – they misinterpret or misassess data on system states and functioning (Woods & Sarter, 2000). Mode error, or confusion about the selected system mode (e.g. Sarter & Woods, 1994; Woods & Sarter, 2000), is a coherence disruption that has resulted in several incidents and accidents. Perhaps the most well-known of these occurred in Strasbourg, France (Ministre de l’Equipement, des Transports et du Tourisme, 1993), when an Airbus-320 confused approach modes:

It is believed that the pilots intended to make an automatic approach using a flight path angle of \(-3.3^\circ\) from the final approach fix . . . The pilots, however, appear to have executed the approach in heading/vertical speed mode instead of track/flight path angle mode. The Flight Control Unit setting of \(\text{\textasciitilde-33}\) yields a vertical descent rate of \(\text{\textasciitilde3300}\) ft/min in this mode, and this is almost precisely the rate of descent the airplane realized until it crashed into mountainous terrain several miles short of the airport (Billings, 1996, p. 178).

The cognitive processes required in high-tech aircraft, then, are quite different than those needed in early days of flying. Achieving and maintaining coherence competence involves data, rationality, logic, and requires the pilot to move toward the analytical modes on the cognitive continuum. This movement affords potential gains as well as potential risks. Analysis in the electronic milieu, as in other arenas, can produce judgments that are much more precise than intuitive judgments. Once the correct landing information is selected and entered into the system, for example, the aircraft can follow an exact three-dimensional path to the runway. However, the analytical process is also much more fragile than intuitive processes – a single small error can be fatal to the process, and one small detail can destroy coherence. Mode confusion, as described in the A-320 accident above, often results from what looks, without accurate analysis, like a coherent picture. The cockpit setup for a flight path angle of \(-3.3^\circ\) in one mode looks very much like the setup for a \(-3300\) ft/min approach in another mode. The proliferation of display characteristics such as this indicates that achieving coherence in the glass cockpit is no easy task. Because analytical coherence is essential within the automated cockpit, the question that needs to be asked is whether cockpit task and display features are conducive to this requirement.
If cognition in today’s automated cockpit is more a matter of analytical coherence than intuitive correspondence, then it is important examine whether the features and properties of the automated cockpit environment elicit the type of cognition required by the flying task. Woods and Sarter (2000), in describing the need for activity-centered system design, underlined the need to discover how “computer-based and other artifacts shape the cognitive and coordinative activities of people in the pursuit of their goals and task context” (p. 339). Hammond et al. (1997), as mentioned, provided evidence that performance depends on the degree to which task properties elicit the most effective cognitive response.

In the electronic cockpit, a discrepancy exists between the type of cognition fostered by current systems/displays and what is required for detection and repair of coherence disruptions. Pictorial displays induce intuition, fostering rapid, imprecise but robust judgments; digital, numerical, or text displays induce analysis, and produce judgments that may be slower and less robust, but are more precise. From the beginning of complex aircraft instrumentation, the trend has been to present data in pictorial, “intuitive” formats whenever possible. For example, the attitude indicator presents a picture of ‘wings’ rather than a number indicating degrees of roll. Collision alert and avoidance warnings involve shapes that change color, and control inputs that will take the aircraft from the “red zone” to the “green zone” on the altimeter. Proposed navigational displays show pilots a series of rectangles forming a “pathway in the sky,” delineating their flight path (Beringer, 2000).

The design and display of most automated systems, then, elicit intuitive or quasi-analytical cognition. Data are pre-processed, and presented in a format that allows, for the most part, a wholistic view of aircraft and system states. Often, pictorial representations exploit human intuitive pattern-matching abilities, and allow quick detection of some out-of-parameter system states. This design philosophy seems to be consistent with the goals of workload reduction and information consolidation – and, indeed, many features of cockpit displays do foster the detection of disruptions to a coherent state. However, current displays may in fact be leading pilots astray by fostering the assumption that all data can be managed in an intuitive fashion. This is a false assumption.

Although pilots can cognitively hover around quasi rationality much of the time if things are operating smoothly, repairing – and often detecting – disruptions to coherence demands a shift toward analysis. Within the seemingly
“intuitive” displays reside numerical data, for example, that signify different commands or values in different modes. When dealing with coherence disruptions, these data must be interpreted – what does this piece of data mean when shown in this color, in this position on the screen, in this flight configuration, in this mode of flight? Once interpreted, data must be compared with expected data to detect discrepancies, and, if they exist, analysis is required to resolve them before they translate into unexpected or undesired aircraft behaviors.

Moreover, before system data can be analyzed, it must first be located. This is often not an easy process, because as the aircraft cockpit has evolved, much of the systems information has either been altered in format or buried below surface displays. The data that would allow for analytic assessment of a situation may not only not be obvious, but not be presented at all or may be buried below surface features. What the pilot sees is an apparently simple display that masks a highly complex combination of features, options, functions, and system couplings that may produce unanticipated, quickly propagating effects if not analyzed and taken into account (Woods, 1996). Woods and Sarter (2000), for example, in describing a typical sequence leading to an “automation surprise,” note that “it seems that the crew generally does not notice their misassessment from the displays of data about the state or activities of the automated systems. The misassessment is detected, and thus the point of surprise is reached, in most cases based on observation of unexpected and sometimes undesirable aircraft behavior” (p. 331). The inability of pilots to track the functioning of cockpit systems is evident in often cited questions they express when describing incidents:

- What is it doing now?
- What will it do next?
- How did I get into this mode?
- Why did it do this?
- I know there is some way to get it to do what I want.
- How do I stop this machine from doing this?
- Unless you stare at it, changes can creep in. (Weiner, 1989; Woods & Sarter, 2000).

Paries and Amalberti (2000) have discussed the dangers of using intuitive, generalized responses when dealing with the “exceptions” that “can be found in the commands or the displays of all existing aircraft” (p. 277). The exceptions they cite involve the logic of the flight mode indications and behavior in relation to the context of the flight – the interpretation of display indications differs depending on the phase of flight and the aircraft. Highly-coupled autopilot
modes make things even more complex. Sherry and his colleagues (Sherry, Feary, Polson & Palmer, 2001; Sherry, Feary, Polson, Mumaw & Palmer, in press) decomposed the functions and displays of the vertical navigation system (VNAV) and of the flight mode annunciator (FMA). They found that the VNAV button is “overloaded” in the descent and approach phases of flight, in that its selection results in the engagement of one of six possible trajectories – and that these trajectories will change autonomously as the situation evolves. Moreover, the same FMA display is used to represent several different trajectories commanded by the VNAV function. It is not surprising, given these sources of confusion, that coherence errors occur. It is impossible to utilize this system intuitively; however, the interface does not provide the necessary information to establish or reinforce correct mental models of system functioning (Sherry et al., 2001).

Complexity and opaqueness also may result in “system-induced misinterpretations,” that is, cases in which crews, in the absence of analytical information, end up making false inferences about the source or nature of a problem and how to resolve it. Plat and Amalberti (2000) observed glass-cockpit crews flying LOFT (Line-Oriented Flight Training) scenarios, and tracked their strategies for handling system faults and failures. Among the principal traits they documented in handling problems for which no procedure was recommended by their electronic assistance system was a “spontaneous tendency to improvise, reset circuit breakers or higher functions, and make (incorrect) inferences anytime the system becomes difficult to understand” (p. 305). Although crews were able to maintain control of the situation, their methods sometimes included “tricks” found by chance, which allowed them to bypass a failure. Plat and Amalberti used the word “magic” several times to describe the thinking and beliefs crews displayed concerning system functions and interactions. Their terminology highlights the need for system and display characteristics that foster correct analytical diagnosis, rather than guesswork.

A mismatch, then, exists between the cognitive tactics demanded to achieve coherence in the automated cockpit and the cognitive tactics elicited by its displays. Automated systems typically analyze information and present only what has been deemed “necessary.” In their efforts to provide an “easy-to-use” display format, designers have often buried the data needed to retrace or follow system actions. Resultant system opaqueness interferes with ability to track processes analytically, a phenomenon that has often been documented and discussed (e.g. Woods, 1996; Sarter, Woods & Billings, 1997). The spontaneous generation or poor mental models of automated system functioning results in mistaken perceptions of coherence, and/or faultily based attempts to achieve and restore it. Clearly, the design of automated systems may hinder rather
than facilitate the cognitive mode needed to maintain coherence in the automated cockpit.

**IMPLICATIONS: TOWARD COHERENCE IN THE AUTOMATED COCKPIT**

Recognition of analytical coherence as a strategic goal in the automated cockpit carries implications for research models, as well as for system design and pilot training. In aviation research, the organizing principles of coherence/correspondence, intuition → analysis can be utilized as theoretical frameworks within which to examine pilot cognition and the cognitive requirements of the automated cockpit, and to ensure a match between them. Within these frameworks, we need to develop new ways of thinking about system design and display features, pilot judgment and decision making, and research paradigms and goals.

*System Design and Display Features*

Woods (1996) referred to automation in electronic cockpits as characterized by “apparent simplicity, real complexity.” His characterization might be modified to describe the cognitive processes demanded by automation in electronic cockpits as “apparently intuitive, really analytical.” Currently, the mismatch between cognitive requirements of the electronic cockpit and the cognitive strategies afforded by systems and displays makes it extremely difficult to achieve and maintain coherence in the cockpit. On one hand, system displays and the opacity of system functioning foster intuition and discourage analysis; on the other hand, the complexity of automated systems makes them impossible to manage intuitively, and requires analytical processing. Recognition of this mismatch is the first step toward rectifying it.

A second step entails acknowledging that the electronic cockpit is a coherence-based, deterministic world. This means that developers have to design systems that are not only reliable in terms of correspondence (empirical accuracy), but are also interpretable in terms of coherence. Principles of “human-centered automation” prescribe that the pilot must be actively involved, adequately informed, and able to monitor and predict the functioning of automated systems (Billings, 1996). To this list should be added the requirement that the design of automation and automated displays elicits the cognition appropriate for accomplishing the human role in the human/automation interaction. Questions that can guide this design requirement include:
• Does the design enable the pilot to understand what cognitive processes are required to utilize a system and how to accomplish goals?
• Does the design facilitate achievement and recognition of coherence and detection of disruptions to it?
• Are the surface and depth characteristics of the display congruent (Hammond et al., 1997) – that is, if depth properties place the task at the analysis-inducing pole of the cognitive continuum, do surface characteristics correspond?
• Does the design facilitate alternation between coherence and correspondence when necessary (e.g. during a visual approach)?
• Do surface features of the system induce the mode of cognition that will enable most effective task performance?

Training for Coherence – Pilot Judgment and Decision Making

When training pilots for the automated cockpit, experiences and expectations should stem from and be tailored to the cognitive parameters of the flying task. Defining and examining cognition in the glass cockpit requires new models of expertise and of judgment and decision making. Correspondence and coherence are separate cognitive processes, and expertise in one does not in any way ensure competence in the other. Models that describe expertise in naturalistic environments, as discussed, focus on intuitive correspondence competence – the use of multiple fallible indicators to recognize and respond to familiar situations – rather than expertise, or coherence competence, in an electronic, deterministic world. These models are complementary to coherence-based models in that they describe critical but different aspects of cognitive behavior. They focus on interaction in an uncertain environment, rather than in a deterministic environment. They are relevant because the electronic world of the cockpit exists within and is subject to the constraints of the natural environment (Vicente, 1990). Situation awareness, for example, entails a knowledge of both correspondence and coherence aspects of a situation – elements outside of the cockpit, such as other aircraft or mountains, as well as within the cockpit, such as autopilot mode or system states (e.g. Endsley, 1996).

Although we have in recent years given much attention in the aviation community to research and training for correspondence expertise, we are lagging behind with respect to improving coherence judgments. Individual differences exist in coherence competence, and selection, training, and practice for it are lacking or uneven across airlines.

Our efforts to facilitate coherence expertise are handicapped, as discussed above, by the mismatch between cognitive requirements for coherence and
features of existent cockpit displays. In the cockpit, coherence requires analysis, in that the data comprising the electronic “story” are impossible to evaluate intuitively and recognitional shortcuts won’t work. Training would therefore need to include instruction and practice in analytical skills relevant to assessing cockpit system functions and displays. Coherence is also very brittle, in that one tiny detail that is the least bit inconsistent (e.g. a decimal point in the wrong place) can destroy it. Training for coherence competence, therefore, would have to overcome these obstacles through the development of highly accurate mental models of how the automation works – ensuring that pilots are able to track the data and information needed to accomplish analysis, and providing an inoculation against the lure of seductive, “intuitive” system displays. Coherence-based critical incidents presented via computer-based trainers or high-fidelity LOFT (Line-Oriented Flight Training) scenarios are potential vehicles for the acquisition and practice of coherence competence.

Training would also need to address the limitations of the electronic world – in particular, situations in which coherence doesn’t achieve correspondence – and conditions under which environmental conditions call for a cognitive switch from coherence to correspondence. Most of the time, in the automated cockpit, coherence assumes correspondence – that is, pilots are led to believe that if the systems present a coherent picture, correspondence accuracy is taken for granted. Given coherence within the cockpit, modern aircraft can be perfectly navigated from A to B, and the pilot never has to look out the window. However, contextual factors and system limitations constrain the accuracy of this assumption, and pilots have to be trained to recognize situations in which contextual (correspondence) factors are critical or are discrepant and must be taken into account. It does the pilot no good, for example, to make a perfect “coherence” landing with all systems in sync if a truck is stalled in the middle of the runway. Moreover, when systems fail drastically and coherence is no longer achievable, correspondence competence may become critical. A dramatic illustration of this type of event occurred in 1989, when a United Airlines DC-10 lost all hydraulics, destroying system coherence and relegating the cockpit crew to intuitive correspondence to manage the aircraft (see Hammond, 2000, for a discussion).

Research Goals

Research paradigms and goals must also be aligned with glass cockpit cognitive requirements. In order to model human behavior in the modern aviation context, it is important to model coherence and correspondence features of the situation, as well as to fine-tune ways to track intuitive and analytical processes.
Speaking, one challenge to the coherence/correspondence, intuition → analysis theoretical framework will be to determine how individual factors, cognitive requirements, and situational features as well as interactions among them can be modeled and assessed so that we can make accurate predictions of human performance. Fortunately, we have the tools in simulation technology to test these models under conditions representative of critical incidents.

Research issues pertaining to these variables and interactions include:

- What is the most efficient and effective combination of correspondence/coherence competence for the pilot of the automated aircraft?
- Under what conditions should pilots switch from coherence to correspondence, and along the continuum from intuition to analysis?
- How do we ensure that pilots practice metacognition – that is, that they maintain awareness of and control over their own cognitive processes?
- If individual differences exist in coherence competence, do certain personality types exhibit a higher resistance to intuition-inducing cockpit displays?
- What are the implications of situational expectations for cognitive processing – what does it take (how much or what sort of an inconsistency), for example, to jolt pilots out of their operational “schema” and perceive a disruption to coherence?
- If analytical coherence is critical in the automated cockpit, how can we ensure sufficient analytical activity without risking cognitive overload?
- Ultimately, as aircraft gain more navigational autonomy under “free flight” rules, what will be the broader relationship between coherence within the cockpit and coherence within the NAS?

Coherence in the electronic cockpit is both attainable and maintainable; however, discovering the most effective means to accomplish this requires a shift in our own cognitive activity, as researchers, toward the models and research paradigms that will allow us to more accurately understand behavior in this milieu. Examining cognition within the correspondence/coherence, intuition → analysis framework offers the possibility of finding solutions across automated cockpit problems. By doing this, we will gain a much broader perspective of the issues and potential problems or hazards within the automated environment, and the ability to make informed predictions regarding the cognitive processes guiding pilot behavior.

**NOTE**

1. Thanks to Ken Hammond for pointing this out at a NASA workshop on human error modeling.
ACKNOWLEDGMENTS

Anyone who is familiar with the work of Ken Hammond will recognize his influence throughout this chapter. Ken is the coherence/correspondence guru, and I am extremely grateful for his help. This work was supported in part by the NASA Aviation Safety Program, NASA Grant # NAG2-1285, and Dr. Judith Orasanu at NASA Ames Research Centre is our Technical Monitor, and has generously given her input and advice.

REFERENCES


6. IDENTIFYING TRAINING AREAS FOR ADVANCED AUTOMATED AIRCRAFT: APPLICATION OF AN INFORMATION-PROCESSING MODEL

Florian Jentsch, James M. Hitt II and Clint Bowers

INTRODUCTION

Recent advances in aircraft instrumentation and equipment have led to a new generation of advanced automated aircraft. The development of electronic flight management systems (FMSs), in particular, has given flight crews very powerful tools to plan, execute, and control specific flight plans by automating navigation and performance calculations. Indeed, the latest generation of FMSs with four-dimensional (4-D) guidance (i.e. along the x-, y-, z-axes, and in time) can conceivably conduct an entire automated flight from engine spool-up at takeoff to deceleration after landing in a fully autonomous way.

While these technical advances have undoubtedly increased the capabilities of modern aircraft, they have also led to a change in the pilot’s job. Previously,
the pilot was largely “in-the-loop,” that is, an active controller who received direct and immediate feedback from the aircraft about the effects of his/her control inputs. With today’s modern flight management systems, however, the pilot’s role has changed from an active controller to a more passive supervisor/monitor/administrator. System inputs made early in the flight, such as route entries, may have adverse consequences much later in the flight. Also, as with other computerized systems, the capabilities of the system are not always entirely transparent to the operator. In fact, recent studies have shown that pilots frequently are uncertain about a system’s status and its future performance (Sarter, Woods & Billings, 1997). Questions such as “What is this thing doing now?,” and “What is it going to do next?” are, unfortunately, frequently heard in today’s modern flight decks.

Sometimes, such uncertainty can have deadly consequences. In the decade since the introduction of the second generation of automated aircraft (Airbus A320, Boeing B757/767), several fatal accidents have occurred in which flight crew uncertainty/error in the operation of advanced autoflight systems has been cited as a causal or contributing factor (e.g. B757 at Cali, Colombia; A320 at Strasbourg, France) (Billings, 1997).

Given the complexity of modern automated aircraft systems and the potential threats that inappropriate flight crew understanding of the systems pose, training for flight crews has become even more important in these aircraft than in traditional aircraft. Already in 1976, Johannsen had pointed out that the changes hitherto brought about by automation had resulted in higher requirements for skills and education of operators. This was also recognized by the U.S. Federal Aviation Administration (FAA) which conducted a systematic review of “glass cockpit” issues (Abbott et al., 1996). Additionally, the FAA sponsored a structured analysis of air carrier needs with respect to automation training which confirmed that the identification of the skills needed to operate automated aircraft successfully is one of the key issues that need to be studied.

As a first step towards defining the skills required for the safe and efficient operation of advanced automated aircraft and towards delineating the strategies, methods, and tools best used to train them, an information-processing (IP) model is applied to the human-automated machine system. The IP model is used to categorize the problems observed in interactions between human operators and advanced automated systems, and to identify those issues that are most likely conducive to improvement through training. Following the identification of potential training areas are brief discussions of the potential efficacy of traditional approaches to performance improvements in these areas.
Operators in many of today’s environments (e.g. aviation, nuclear power, maritime, medicine) must interact with complex automated systems to perform their tasks. These systems are designed to automatically execute a wide variety of actions that have been traditionally performed by human operators (e.g. monitoring, control, and communication of system status). As a result, these systems often exhibit high levels of complexity, autonomy, and authority. Despite considerable advancements in the design of these systems, performance is not always optimal (Billings, 1997; Lee & Sanquist, 1996; Sarter & Woods, 1994).

A number of problems and failures regarding the interaction between humans and advanced automated systems have been identified, such as failures to detect system changes and system malfunctions, failures to identify and classify system problems, failures to maintain situation awareness, and failures to appropriately use the system to achieve specific tasks. Traditionally, many of these failures have been studied in isolation and from a limited perspective, such as system design or human performance modeling (e.g. Abbott et al., 1996). For example, a large body of research has investigated attention-demands and -problems in automation systems (cf. Parasuraman, Mouloua, Molloy & Hilburn, 1996). While this research has contributed significantly to our understanding of human limitations in interactions with advanced automated systems, it has provided a rather limited set of information for training.

The current chapter focuses on developing enhanced training methods and tools that can be implemented in air carrier training environments, on testing research hypotheses, and on generating principles and guidelines for the design of aircrew automation skill training. Towards this end, an information-processing model was used to categorize the problems observed in interactions between human operators and advanced automated systems, and to identify those issues that are most likely conducive to improvement through training. Following the identification of potential training areas are brief discussions of traditional approaches to performance improvements in these areas.

Introduction of the Information-Processing Model

The use of human performance models or taxonomies to analyze potential and observed problems of human-machine interactions has a long history in many areas
of human factors. Occasionally, models have also been used in automation research. For example, Kantowitz and Campbell (1996) used a Human Information Processing (HIP) taxonomy to analyze incident reports involving automation to study the effects of automation on workload. Others have used task-models to classify accidents and to derive design solutions to observed problems.

The information-processing (IP) model selected for this project was proposed by Wickens (1992). It is shown in Fig. 1. While there are other information-processing models in the literature, Wickens’ IP model is useful for this project as it provides a systematic integration of major constructs in human-factors psychology, such as sensation and perception, attention, memory, decision making, etc. Also, Wickens’ IP model provides a step-by-step guide for the analysis of human-machine interactions. For each human-machine interaction, system characteristics can be compared against known human performance limitations. For example, display characteristics can be compared against human limitations in sensation (e.g. is a signal bright enough to be picked up by the human sensory system?), perception (e.g. can a signal be perceived as different from any existing background noise?), attention (e.g. what are the attention requirements to detect this signal and do they match the human capabilities as

![Fig. 1. Human Information Processing (IP) Model. Adapted from Wickens (1992).]
the task is being executed?), memory (e.g. is the signal one of many similar ones that have to be learnt and then compared in working memory?), etc. By using the IP model as a checklist/guide, it is therefore possible to identify and classify potential or existing limitations/problems in the human-machine system interaction.

Description of the Model

As indicated in Fig. 1, Wickens’ IP model is a closed-loop system with discrete processing stages and resources (Wickens, 1992; Wickens & Flach, 1988). Stimuli enter the closed-loop system and, if of sufficient strength, are processed by the sensory system. Then, these stimuli can become percepts and are coded based on physical attributes such as intensity, pitch, tone for auditory sounds and brightness, luminance, color, and shape for visual stimuli – provided they are attended to. The processed information then enters what is called the decision and response selection stage. It is here where the coded percepts enter working memory. Unless rehearsed or encoded into long term memory, the percepts quickly cease to exist. But if an action is chosen, the information moves along into the response execution stage where motor action commences. From these motor responses, new stimuli are created and the process begins over again creating a closed-loop.

One of the keys to this model is the use of attentional resources. Attentional resources must be used in the perceptual, cognitive (decision and response selection), and response execution stages. There is a limited amount of total attention one person can exhibit during this process. If at any time these attentional resources are depleted, the process breaks down at that moment and processing becomes incomplete.

Four major premises follow from Wickens’ IP model:

(1) Human information processing is based on the existence of discrete resources for sensation, perception, memory, decision making and execution, and attention.
(2) The discrete resources are applied in a sequentially interdependent, closed-loop fashion (i.e. the output of a preceding stage becomes the necessary input for a subsequent one, with environmental conditions providing the feedback to the information-processor’s actions).
(3) The discrete resources are limited, specifically as attention is concerned.
(4) Potential problems in human-machine interactions can be identified by comparing the characteristics of the task situation/environment to the human limitations at each discrete resource.
THE DEFINITION OF AUTOMATION AND ITS CONFOUNDING WITH OTHER CONSTRUCTS

One problem frequently encountered when studying problems of human-machine interaction in highly advanced or automated system is the confounding of automation with other technological advances. As pointed out by Bowers, Jentsch, and Salas (1995), the introduction of automated functions often occurs simultaneously with changes in areas such as information display, control technology, etc. Indeed, the term “glass cockpit” that is commonly used to describe the cockpit of advanced, automated aircraft, more accurately describes the new display technologies (i.e. cathode ray tubes [CRTs] or liquid-crystal displays [LCDs]) employed in these aircraft, rather than the automation of guidance and control functions. As Woods (1996) put it, “automation is a wrapped package – a package that consists of changes on many different dimensions bundled together as a hardware/software system” (p. 4).

Consequently, many problems observed in advanced automated aircraft may be the result of technological advances other than the automation of functions, e.g. display systems, or may be due to the integration of other advanced technologies with automation. Woods and Sarter (1993, p. 107), for example, pointed out that . . .

... problems are not inherent in automation per se but are rather the result of a clumsy design and implementation of intelligent systems. For example, advanced technological capabilities such as color graphic displays and multiple window interfaces are often used for decorative purposes rather than for supporting the coordination and communication between all agents in the system.

Even in cases, however, where changes in displays and controls are not merely introduced for decorative or “design” purposes, automation often can only achieve its intended results through concurrent changes in the human-machine interface (Johannsen, 1976). Bauerschmidt and LaPorte (1976), for example, suggested that the requirements for operator involvement in actual control tend to be reduced as the displays and controls are increasingly integrated. This is further demonstrated in the following section.

Automation of One Function Often Necessitates Other Changes

While the appropriate allocation of functions to humans and machines has long been the driving force behind automation, adding automation is more than a transfer of one specific function from the human to the automated system. Instead, it often requires or results in other changes. Sarter, Woods and Billings (1997, p. 1931) pointed out:
When new automation is introduced into a system or when there is an increase in the autonomy of automated systems, developers often assume that adding “automation” is a simple substitution of a machine activity for human activity (the substitution myth). Empirical data on the relationship of people and technology suggest that is not the case. Instead, adding or expanding the machine’s role changes the cooperative architecture, changing the human’s role, often in profound ways.

The following example shall be used to illustrate point by Sarter et al. (1997) that the addition of any machine tool of function can result in a substantial role change for the human operator. Hutchins (1995) analyzed the task structures and allocations for identifying and maintaining the correct airspeeds during approach. Hutchins’ example demonstrated how even simple mechanical (i.e. non-automated) devices can fundamentally change the role of human operators (Hutchins, 1995, p. 283):

Without a speed bug, on final approach the PF [i.e. pilot flying] must remember the approach speed, read the airspeed indicator scale, and compare the position of the ASI [i.e. airspeed indicator] needle on the scale with the position of the approach speed on the scale. With the salmon bug set, the pilot no longer needs to read the airspeed indicator scale. He or she simply looks to see whether or not the indicator needle is lined up with the salmon bug. Thus, a memory and scale reading task is transformed into a judgment of spatial adjacency.

If even a simple mechanical device such as a speed bug can alter the human role in such fundamental ways, it is not surprising that the actual “automation” of functions can have even more profound effects. However, Hutchins’ (1995) example of speed bugs also demonstrates how the automation of functions often must be partitioned from corresponding changes in the displays and controls which form the interface between the human operator and the (automated) machine. Automating the approach speed control task, for example, can take place at different levels. At one level, one may want to automate the calculation of the target speeds from the actual aircraft weight; this task is traditionally achieved by the pilots through the use of tables or charts. Once the speeds are calculated, they must be displayed somehow for the pilots. Likewise, one may want to automate the subtask of setting the speeds into some form of memory aid (i.e. the speed bugs). Speed bugs may be set manually for the calculated target speeds, or they may be set automatically. Finally, holding the speeds and extending the gear and appropriate lift devices (i.e. flaps and slats) can take place manually or automatically. In each case, displays are needed to convey the actual status of the situation for the pilots, unless one believes in fully autonomous operation of the automated system, in which case specific displays may not be needed – but might still be helpful. All this complexity then, of task allocation, functional change, etc., almost invariably results in changes in
automation that, frequently, are hopelessly confounded with changes in the associated displays and controls.

Automation Defined

Given the invariable confounding of advances in automating functions with those in other areas of technology, it is perhaps useful at this point to review “automation” as a construct in more detail. “Automation” in its strictest sense refers to the transfer of functions from a human operator to a machine. This definition, however, is very broad, not making it very diagnostic (Wiener, 1988). Specifically, it has frequently been pointed out that automation is not a dichotomy (i.e. off or on; see Kantowitz & Campbell, 1996), and that it is important to distinguish between what is automated, for whom, and how (Jentsch & Bowers, 1996).

Despite this observation, however, it is surprising how the majority of the literature on various aspects of the interactions between humans and advanced, automated system discusses “automation” as if it were a unified, well-defined, and singular concept. As a consequence, all sorts of effects and problems are attributed to “automation” in general, even if these effects or problems were only observed under very specific circumstances and in very specific conditions of automation. In our view, it is therefore very important to review a number of frameworks that have been proposed to define “automation” in more detail.

Control vs. information vs. management automation. Wiener (1988) distinguished automation according to two dimensions: Automation of control functions and automation of monitoring functions. According to Wiener, automation of control functions relates to issues “such as flight path guidance, power plant, or environmental control” (p. 436). Automation of monitoring functions, on the other hand, relates to a form of machine monitoring and alerting, such as the MD-80s central aural warning system (CAWS) (Wiener, 1988). In automation of monitoring functions, the machine monitors system parameters and advises the human operator about the status and in the case of exceedances.

Billings (1997) expanded on this work by Wiener and by others, such as Fadden (1990), by providing a similar functional classification that distinguishes between three categories of automation: In addition to control automation and information automation (similar to automation of monitoring functions in Wiener’s, 1988, model), Billings added a third functional form of automation, which he called “management automation” (Billings, 1997, p. 70):
Table 1. Levels of automation (adapted from Billings, 1997).

<table>
<thead>
<tr>
<th>Management Mode</th>
<th>Automation Functions</th>
<th>Human Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Autonomous Operation</strong></td>
<td>Fully autonomous operation; Pilot not usually informed; System may or may not be</td>
<td>Pilot generally has no role in operation</td>
</tr>
<tr>
<td></td>
<td>capable of being disabled.</td>
<td>Monitoring is limited to fault detection</td>
</tr>
<tr>
<td></td>
<td>Pilot informed of system intent;</td>
<td>Goals are self-defined; pilot normally has no reason to intervene.</td>
</tr>
<tr>
<td></td>
<td>Must consent to critical decisions;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>System informs pilot and monitors responses.</td>
<td></td>
</tr>
<tr>
<td><strong>Management by Exception</strong></td>
<td>Essentially autonomous operation; Automatic reconfiguration; System informs pilot and monitors responses.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Full automatic control of aircraft and flight.</td>
<td>Pilot must consent to state changes, checklist execution, anomaly resolution;</td>
</tr>
<tr>
<td></td>
<td>Intent, diagnostic and prompting functions provided.</td>
<td>Manual execution of critical actions.</td>
</tr>
<tr>
<td><strong>Management by Delegation</strong></td>
<td>Autopilot &amp; autothrottle control of aircraft and flight.</td>
<td>Pilot commands heading, altitude, airspeed;</td>
</tr>
<tr>
<td></td>
<td>Automatic communications and nav following.</td>
<td>Manual or coupled navigation;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commands system operations, checklists, communications.</td>
</tr>
<tr>
<td><strong>Shared Control</strong></td>
<td>Enhanced control and guidance; Smart advisory systems;</td>
<td>Pilot in control through control-wheel steering or envelope-protected system;</td>
</tr>
<tr>
<td></td>
<td>Potential flight path and other predictor displays.</td>
<td>May utilize advisory systems;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>System management is manual.</td>
</tr>
<tr>
<td><strong>Assisted Manual Control</strong></td>
<td>Flight director, Flight Management Systems, navigation modules; Data link with manual messages; Monitoring of flight path control and aircraft systems.</td>
<td>Direct authority over all systems;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manual control, aided by Flight director and enhanced navigation displays;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flight Management System is available;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trend info on request.</td>
</tr>
</tbody>
</table>
When management automation is available, the pilot has the option of acting as a supervisory controller (Sheridan, 1997). In aircraft, automation is directed by the pilot to accomplish the tactical control functions necessary to accomplish the objective. Using this classification schema, an autopilot altitude hold function would be a feature of control automation, whereas the Engine Information and Control System (EICAS) would mostly fulfill information automation functions. Finally, advanced Flight Management System (FMS) may allow some degree of management automation.

**Levels of automation and authority.** A second way to distinguish different forms of automation was proposed early by Sheridan (1976) and has since been expanded upon by others (e.g. Billings, 1997; Rouse & Rouse, 1983). The models/conceptualizations in this class do not distinguish between “what is automated.” Instead, they focus on “how automated it is,” i.e. the degree to which the human operator has influence on the functions of the (automated) system and how tasks are partitioned between the operator and the machine. A commonality among all models is that they describe a continuum with total manual control and operation on one end and complete automation at the other. The models differ only in the number of levels they propose (i.e. how finely the continuum from manual to automatic is sliced). Table 1 shows Billings’ (1997) conceptualization of this type of framework for pilots with seven levels of automation.

**Summary of automation definitions.** Together, the two frameworks for the definition of automation discussed here can be used to classify human-automated machine system according to several dimensions: “What” is automated (i.e. a functional classification), and “to what degree” it is automated. Added to this in a team context could be “whose” function in a multi-pilot crew is automated.

<table>
<thead>
<tr>
<th>Management Mode</th>
<th>Automation Functions</th>
<th>Human Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Manual Control</strong></td>
<td>Normal warning and alerts; Voice communication with Air Traffic Control; Routine ARINC Communication and Radio System communications performed automatically.</td>
<td>Direct authority over all systems; Manual control using raw data; Unaided decision-making; Manual communications.</td>
</tr>
</tbody>
</table>
Using such a framework of “what,” “how much,” and “whose” functions are automated, Jentsch and Bowers (1996) were able to show vastly different outcome effects of “automation” on crew performance. In their study, the automation of a simple control function (i.e. altitude hold) at the level of management by consent for the pilot-flying, for example, freed up the pilot-flying from a workload-intensive control task. This resulted in a reduction of workload, increases in crew communications, and better team performance on a secondary monitoring task. Conversely, automation of an information function (i.e. calculation and display of position coordinates) at the level of assisted manual control for the pilot-not-flying, resulted in increased needs for heads-down time for the pilot-not-flying while he/she attempted to control and read the display computer. This, in turn, was associated with an increase in attention demands, a simultaneous reduction in mental workload, a reduction in crew communication, and unchanged task performance in monitoring.

Jentsch and Bowers (1996) were able to use these results to reconcile a number of heretofore irreconcilable observations from the study of crew coordination in automated aircraft, where some studies had shown increases in crew coordination during “automated” flight, while others had shown corresponding reductions. In each case, the type, level, and target of the automation had been different, resulting in vastly different results that became entirely plausible once a more specific approach to the definition of automation was taken.

**Conclusion**

A consequence of the confounding of automation with other advanced technologies is that problems in advanced systems may be the result of changes either in automation or in other technologies, or may be the result of interactions between automation and one (or several) other technologies. Sometimes, it may be difficult to decompose which construct is responsible for a particular problem, especially when the analysis is outcome- rather than process-driven and does not take the various types, levels, and targets of automation into account. The use of a systematic model such as Wickens’ IP model, however, greatly facilitates the analysis and allows the development of a more detailed, accurate picture. For example, using the IP model, a specific problem with an advanced display may be identified in the analysis of the sensory and perceptual demands as a mismatch between display characteristics and the human sensory system. In contrast, another apparent display problem may be more correctly identified as relating to the impact of automation on attention. Consequently, applying a systematic model in the analysis of observed problems can add significant detail to the development of appropriate solutions.
APPLICATION OF WICKENS’ IP MODEL TO OBSERVED PROBLEMS IN ADVANCED SYSTEMS

Listing of Known Problems in Interactions between Humans and Automated Aircraft Systems

As advanced automated systems have become more prevalent, a number of studies have investigated existing problems in the interactions of human operators with these systems. One structured listing of problems in interactions between humans and automated systems, and specifically those systems in aviation, has been provided by Funk and Lyall (1997). Based on accidents and incidents, interviews, research reviews, and questionnaire data, Funk and Lyall identified a set of approximately 90 issues that relate to the interaction of humans with automated systems.

Of the 92 human factors problems associated with flightdeck automation, 61 related to training for automated aircraft. Table 2 contains a list of the 61 issues that were of direct interest. These issues were applied to the analysis via the IP model to identify areas for potential training interventions. Each observed problem could be categorized as relating to one or several of the parts in the IP model (see Table 3). This analysis indicated that the observed, larger automation issues were relatively equally distributed among the functional parts of the IP model, with a slight preponderance of issues in the areas of memory and decision making. Another area that showed a high number of problems was sensation, mostly as the display of data was concerned. This may indicate that in many cases, “automation-induced” problems were related more to accompanying changes in display technology, and less to actual automation effects. Finally, attention and, as a secondary issue not directly shown in Wickens’ model, workload (which could be understood as an issue related to the limited resource of attention) were cited numerous times in the problem statements.

Applying the IP Model: Accidents and Incidents

While scientific analyses of problems between humans and advanced automated aircraft systems, such as the analysis by Funk and Lyall (1997), often report general issues rather than detail problems, the analysis of raw data, e.g. in the form of accident or incident reports, can shed some additional light on the multi-faceted nature of interaction problems between humans and machines. The following examples will demonstrate the utility of the IP model in these instances, using a selection of accidents and incidents involving advanced civilian aircraft.
**Table 2.** List of Relevant Automation Issues from the Funk and Lyall (1997) Database.

<table>
<thead>
<tr>
<th>Database Issue No.</th>
<th>Description of Automation Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
<td>pilots may be out of the loop</td>
</tr>
<tr>
<td>005</td>
<td>monitoring requirements may be excessive</td>
</tr>
<tr>
<td>007</td>
<td>manual skills may not be acquired</td>
</tr>
<tr>
<td>009</td>
<td>information integration may be required</td>
</tr>
<tr>
<td>011</td>
<td>automation integration may be poor</td>
</tr>
<tr>
<td>014</td>
<td>information overload may exist</td>
</tr>
<tr>
<td>023</td>
<td>failure recovery may be difficult</td>
</tr>
<tr>
<td>024</td>
<td>failure modes may be unanticipated by designers</td>
</tr>
<tr>
<td>025</td>
<td>failure assessment may be difficult</td>
</tr>
<tr>
<td>026</td>
<td>pilots may be reluctant to assume control</td>
</tr>
<tr>
<td>037</td>
<td>controls of automation may be poorly designed</td>
</tr>
<tr>
<td>038</td>
<td>scan pattern may change</td>
</tr>
<tr>
<td>039</td>
<td>interface may be poorly designed</td>
</tr>
<tr>
<td>040</td>
<td>automation may be too complex</td>
</tr>
<tr>
<td>044</td>
<td>mode transitions may be uncommanded</td>
</tr>
<tr>
<td>046</td>
<td>pilots may lack confidence in automation</td>
</tr>
<tr>
<td>047</td>
<td>data access may be difficult</td>
</tr>
<tr>
<td>049</td>
<td>data re-entry may be required</td>
</tr>
<tr>
<td>053</td>
<td>vertical profile visualization may be difficult</td>
</tr>
<tr>
<td>055</td>
<td>manual operation may be difficult after transition from automated control</td>
</tr>
<tr>
<td>063</td>
<td>deficiencies in basic aircraft training may exist</td>
</tr>
<tr>
<td>065</td>
<td>manual skills may be lost</td>
</tr>
<tr>
<td>070</td>
<td>false alarms may be frequent</td>
</tr>
<tr>
<td>071</td>
<td>data entry errors on keyboards may occur</td>
</tr>
<tr>
<td>072</td>
<td>cross checking may be difficult</td>
</tr>
<tr>
<td>075</td>
<td>both pilots’ attention simultaneously diverted by programming</td>
</tr>
<tr>
<td>083</td>
<td>behavior of automation may not be apparent</td>
</tr>
<tr>
<td>087</td>
<td>data presentation may be too abstract</td>
</tr>
<tr>
<td>089</td>
<td>new tasks and errors may exist</td>
</tr>
<tr>
<td>092</td>
<td>displays (visual and aural) may be poorly designed</td>
</tr>
<tr>
<td>095</td>
<td>mode awareness may be lacking</td>
</tr>
<tr>
<td>099</td>
<td>insufficient information may be displayed</td>
</tr>
<tr>
<td>101</td>
<td>automation use philosophy may be lacking</td>
</tr>
<tr>
<td>103</td>
<td>automation level decisions may be difficult</td>
</tr>
<tr>
<td>105</td>
<td>understanding of automation may be inadequate</td>
</tr>
<tr>
<td>106</td>
<td>pilots may over-rely on automation</td>
</tr>
<tr>
<td>107</td>
<td>workarounds may be necessary</td>
</tr>
<tr>
<td>108</td>
<td>automation behavior may be unexpected and unexplained</td>
</tr>
<tr>
<td>109</td>
<td>automation may lack reasonable functionality</td>
</tr>
<tr>
<td>112</td>
<td>data entry and programming may be difficult and time consuming</td>
</tr>
<tr>
<td>114</td>
<td>situation awareness may be reduced</td>
</tr>
<tr>
<td>117</td>
<td>function allocation may be difficult</td>
</tr>
</tbody>
</table>
Example No. 1 – China Airlines A-300-600R, Nagoya, Japan, 1994. One example illustrating the usefulness of applying the IP model to problem in automated systems is the accident in 1994 of an A-300-600R which killed 264 on board, sparing only 7 passengers. The accident report summary indicated that the aircraft was inadvertently placed in take-off go-around (TOGA) mode during a normal approach to runway 34 under Visual Meteorological Conditions (VMC). The captain warned the co-pilot (who was also the pilot at the controls; or “Pilot Flying” – PF) of the mode change three times within 30 seconds of activation. Despite these warnings, and the added power and pitch-up of the aircraft, the PF continued to keep the aircraft on the glideslope. As the pitch-up continued, the PF attempted to manually drop the nose but was met with great resistance from the control yoke. The aircraft pitching increased in excess of 50° at which time the plane stalled, and descended uncontrolled until crashing the ground tail-first. While one report (Sekigawa & Mecham, 1996) indicated automation-related problems were probable causes in the accident investigation, other reports were more critical of the pilots’ behavior. Supporting this notion was the publication of several NTSB Safety Recommendations (A-94-164 through 166) stating, “. . . the accident in Nagoya and the incident in Moscow [another accident, the author] indicate that pilots may not be aware that under

Table 2. Continued.

<table>
<thead>
<tr>
<th>Database Issue No.</th>
<th>Description of Automation Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>operational knowledge may be lacking in design</td>
</tr>
<tr>
<td>122</td>
<td>automation may use different control strategies than pilots</td>
</tr>
<tr>
<td>123</td>
<td>inadvertent autopilot disengagement may be too easy</td>
</tr>
<tr>
<td>128</td>
<td>complex automation may have overly simplistic interface</td>
</tr>
<tr>
<td>129</td>
<td>transitioning between aircraft may increase training requirements</td>
</tr>
<tr>
<td>130</td>
<td>transitioning between aircraft may increase errors</td>
</tr>
<tr>
<td>131</td>
<td>pilots may be overconfident in automation</td>
</tr>
<tr>
<td>132</td>
<td>training may be inadequate</td>
</tr>
<tr>
<td>138</td>
<td>standardization may be lacking</td>
</tr>
<tr>
<td>139</td>
<td>inter-pilot communication may be reduced</td>
</tr>
<tr>
<td>140</td>
<td>automation information in manuals may be inadequate</td>
</tr>
<tr>
<td>145</td>
<td>mode selection may be incorrect</td>
</tr>
<tr>
<td>146</td>
<td>pilots may under-rely on automation</td>
</tr>
<tr>
<td>149</td>
<td>automation may not work well under unusual conditions</td>
</tr>
<tr>
<td>151</td>
<td>procedures may assume automation</td>
</tr>
<tr>
<td>152</td>
<td>state prediction may be lacking</td>
</tr>
<tr>
<td>158</td>
<td>planning requirements may be increased</td>
</tr>
<tr>
<td>167</td>
<td>task management may be more difficult</td>
</tr>
</tbody>
</table>
Table 3. Categorization of Relevant Automation Issues from the Funk and Lyall (1997) Database to Constructs from the Information-Processing (IP) Model.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Issue No.s from the database</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensation:</strong></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>009, 011, 039, 047, 087, 092, 099, 138, 139, 152</td>
</tr>
<tr>
<td><strong>Perception:</strong></td>
<td></td>
</tr>
<tr>
<td>Pattern matching</td>
<td>025, 070, 103</td>
</tr>
<tr>
<td><strong>Memory &amp; Decision and Response Selection:</strong></td>
<td></td>
</tr>
<tr>
<td>WM: Dynamic world model (including SA)</td>
<td>002, 053, 083, 095, 114, 128, 158</td>
</tr>
<tr>
<td>LTM: Mental model</td>
<td>024, 108, 122, 145</td>
</tr>
<tr>
<td>LTM: Knowledge</td>
<td>040, 044, 063, 089, 101, 105, 117, 121, 129, 130, 133, 140, 151</td>
</tr>
<tr>
<td>LTM: Experience and exposure</td>
<td>007, 026, 038, 046, 106, 109, 146, 147, 150</td>
</tr>
<tr>
<td><strong>Response Execution/Motor:</strong></td>
<td></td>
</tr>
<tr>
<td>Actions (motor)</td>
<td>026, 037</td>
</tr>
<tr>
<td>Movements (outcomes)</td>
<td>049, 055, 065, 071, 112, 123</td>
</tr>
<tr>
<td><strong>Attentional Resources:</strong></td>
<td></td>
</tr>
<tr>
<td>Resources of attention</td>
<td>005, 023, 072, 075, 131</td>
</tr>
<tr>
<td>Goals</td>
<td>107</td>
</tr>
<tr>
<td>Workload</td>
<td>014</td>
</tr>
<tr>
<td>Issues which mention increased workload as a possible result of automation problem</td>
<td>009, 011, 014, 024, 047, 053, 103, 108, 109, 121, 138, 151</td>
</tr>
</tbody>
</table>
some circumstances the autopilot will work against them if they try to manually control the airplane (p. 5)” (Billings, 1997, pp. 318–319).

Using this accident and the safety recommendations published by the NTSB one can see how the IP framework can be utilized to diagnose some of the various breakdowns of the human-computer interaction. The behavior of the pilot flying would indicate an interaction problem falling within the category of “Memory & Decision” (both long term memory and experience and exposure) and “Response Selection.”

**Example No. 2 – Delta Air Lines McDonnell-Douglas DC-9, Boston, 1973.** A second example for the usefulness of applying the IP model to problems in automated systems is the following accident which occurred in July 1973 and resulted in all 89 on board being killed (Gero, 1996). Although the accident involved a McDonnell-Douglas DC-9, a conventional aircraft with no advanced automation, the aircraft did have some automated functions in the form of an autopilot and a flight director. The accident illustrated an early problem of mode awareness with these systems. The information about the accident was related by Billings (1997, p. 75):

... A Delta DC-9 impacted a seawall short of a runway at Boston; its crew is believed to have followed the flight director, which was miss set in “attitude” rather than “approach” mode, without adequate cross-checking of localizer and glide slope.

While the source of the problem in this accident may at first appear to have been the flight director display, application of the IP model quickly demonstrates that there were no mismatches between the display and human sensory or perceptual capabilities. Likewise, information automation itself was appropriately working in that the commands displayed on the flight director were correctly calculated commands for the attitude mode. Thus, while the automation worked as intended and the pilots were effectively following the calculated commands, what the display was showing was not what the pilots needed and expected at that time. Thus, the mode problem identified here had to do with an attention problem in management automation (i.e. identifying what to display when, and which data the display was using for its commands).

**Example No. 3 – British Midlands Boeing B-737, Birmingham, 1989.** A final incident shall be used to illustrate the utility of the IP model for accident and incident analysis in advanced aircraft. In 1989, a Boeing B-737 crashed only 0.6 NM short of the East Midlands Airport near Birmingham, England (Department of Transport, 1989; Job, 1996). Earlier in the flight, the crew had experienced an intermittent vibration in one of the engines and some smoke in the cockpit. In response, the pilots decided to conduct a precautionary engine shutdown, single-engine diversion, and landing. Unfortunately, the crew shut
down the wrong engine, and the second engine failed shortly before reaching the airport, resulting in a crash that killed 47 and injured 59 of the 111 on board. The following findings from the British accident report are related to automation use and were provided by McClumpha and Rudisill (1994, p. 290):

... some time was lost as the copilot attempted unsuccessfully to program the flight management system (FMS) to produce the correct flight instrument display for landing at East Midlands Airport. From the cockpit voice recorder, it is inferred that the first officer selected the route page and was entering the correct information (for the destination airfield) but as an enroute point. He did not notice the inadvertent selection nor understand the limitations of the available selections with respect to this information. An absence of appropriate feedback within the FMS allowed the error to remain. The first officer failed to select the arrival airfield page; this page is similar to the enroute data page in terms of data layout and data entry.

Two areas of the IP model are implicated in the description of this accident. First, the similarity of the interface for the two modes made it difficult for the pilot, who was under considerable stress, to identify which one he was using, indicating a perception-attention problem. Second, the description of the accident indicates that the pilot never developed the correct knowledge with respect to the system capabilities in the different modes. Using the IP model, one would identify this as a problem with the development of correct mental models or knowledge structures in long-term memory.

**Comparison with Another Automated System: Analysis of the FAA STARS System**

The Standard Terminal Automation Replacement System (STARS) was an air traffic control (ATC) computer-, display-, and control-system that was developed to replace the Automated Radar Terminal Systems (ARTS) heretofore used at airports throughout the United States. STARS required the modernization and replacement of major hard- and software in the U.S. ATC system. Specifically, the STARS program was to achieve the following phased goals which included the automation of major system functions (FAA, 1997, p. 1):

... replaces the current ARTS display consoles and the Digital Bright Radar Tower Equipment with new display hardware while maintaining the Existing Automation Service Level (EASL). It also provides a controller interface to the new Emergency Service Level (ESL) back-up system. A new Monitor and Control Workstation (MCW) also will be implemented for Airway Facilities (AF) personnel at the TRACONs. The [next] stage replaces the ARTS computers with new central computers for radar and flight data processing. It also provides the infrastructure needed to support interfaces to new Air Traffic (AT) service applications and to the enhanced Traffic Flow Management (TFM) system. In the [final] stage, new functions will be implemented for controllers. These include a range
of automation capabilities that are currently in operational use at field facilities or under research and development by several government agencies.

As indicated in the preceding quote, STARS in its final form, was to be a highly automated human-machine system that changed not only the level of automation for a number of tasks, but also made significant changes to the display- and control-technology used in the terminal area air traffic control. As such, the STARS system was comparable to many advanced flightdeck systems which were developed with similar intention, within similar timeframes, and by using similar technologies.

In 1997, a systematic evaluation of the human factors issues involved in STARS identified a number of problems with the system (FAA, 1997; Mogford et al., in press; STARS, 1998). Because of the advanced nature of the system and the human-factors focus of the study, the findings from this review lend themselves for a comparison with the problems and issues identified in the automated cockpit. In the initial human factors evaluation of the new STARS display and control console (i.e. the Monitor and Control Workstation, or MCW), 89 human factors issues were identified. Additionally, a number of other problems were identified during a second review. Together, this resulted in 113 issues related to the system and its display- and control-interfaces.

Using Wickens’ IP model as the guideline for the classification of these issues, a categorization of the observed human factors problems can be seen in Table 4. Also listed are the same categorizations we had made earlier among the issues in the database of automation issues by Funk and Lyall (1997).

Examining Table 4, several statements can be made. The percentages for each category (except Perception) are very similar, showing the similarity in problems for the two systems. The STARS evaluation noted that many of the major criticisms of the system were related to perceptual issues such as color coding, fonts, tabular displays, and consistency of displays (STARS, 1998). Of particular importance for our purposes is the large amount of problems classified under the “Memory & Decision and Response Selection” and “Attentional Resources” categories. Both of these categories had the highest percentages of HCI problems for both systems. While this classification is not very detailed, it does supply the researchers with an idea of areas which need focus in future designs and training requirements.

**SUMMARY AND CONCLUSIONS**

The application of Wickens’ (1992) information-processing (IP) model to an array of automation-related issues, incidents, and systems consistently identified
a number of problems related to the interaction of human users/operators and advanced, automated systems. Given the nature of the IP model, these problems and issues clustered into five major groups, roughly along the demarcations of the major components of the IP model (i.e. “sensation & perception,” “memory,” “decision making,” “attention,” and “decision execution”). The five groups of problems were:

(1) **Problems related to mismatches between display hard- and software and the capabilities of the human sensory/perceptual system.** A large number of problems identified from the study of automated systems was actually unrelated to the automation of system functions or actions as such. Instead, these were problems related to the display of information, whether derived from an automated system or not. The (inappropriate) use of color to code traffic information, for example, is relatively independent of the degree of automation (recall the definition discussed earlier) of the tasks (i.e. traffic avoidance) that the display is used for. McClumpha and Rudisill (1994) similarly reported that “one set of comments [by pilots about civil flight deck automation] related to specific systems and specific problems with the HCI [i.e. human computer interaction] on the aircraft” (p. 290).

(2) **Problems related to increased requirements for vigilance, attention-sharing, and – distribution.** The second group of problems commonly identified in the analysis of interactions between humans and machines was concerned with the constructs of vigilance, attention-sharing, and attention distribution. With the introduction of automation, the need for increased vigilance has become higher than ever as the operator has become both an active (manual

---

**Table 4.** Comparison of Human Factors Issues in the STARS Review and in the Funk & Lyall Database.

<table>
<thead>
<tr>
<th></th>
<th>STARS (113 issues)</th>
<th>Funk &amp; Lyall (61 issues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensation</td>
<td>10.7%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Perception</td>
<td>32.1%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Memory &amp; Decision and Response</td>
<td>79.5%</td>
<td>54.1%</td>
</tr>
<tr>
<td>Selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attentional Resources</td>
<td>33.0%</td>
<td>31.1%</td>
</tr>
<tr>
<td>Response Execution/Motor</td>
<td>8.9%</td>
<td>13.1%</td>
</tr>
</tbody>
</table>

*Note:* Table 4. Percentages of HCI problems categorized for two different systems (ATC and flightdeck) using the IP framework. Note: Each HCI problem can be classified under more than one category. This results in the columns not adding up to 100%.
operation) and a passive (monitoring an automated system) agent in the human-machine system. One common example illustrating the importance of increased vigilance, attention-sharing, and attention distribution occurred under a high workload condition in the A320 crash at Strasbourg in 1992. The pilots received radar vectors leaving them little time for FMS setup for a VOR-DME approach. It is believed the pilots were attempting to program the aircraft for an automatic approach using a flight angle of –3.3 degrees. While believing to execute this command, the aircraft was actually still in the heading/vertical speed mode and not the desired track/flight path angle mode. Therefore, the aircraft began an automatic descent at a rate of –3300 ft/min and crashed into the surrounding mountains (Billings, 1997).

A lack of necessary vigilance towards the current state (mode) of the automation combined with attentional resources being divided between preparation for landing, avoiding air traffic, communicating with ATC, and programming the automation must be viewed as possible breakdowns in the information processing of the crew members.

(3) Problems related to the development of accurate knowledge in long-term memory. A third group of problems in automated systems has to do with shortcomings that operators have in their knowledge and understanding of automated systems. Pilots, for example, often do not leave training with the appropriate and accurate knowledge in long-term memory (Amalberti, 1999, p. 178):

Most pilots lack a fundamental grasp of the internal logic of automation and evaluate the gap between their expectations (governed by what they would do if they were in command) and what the computer does. This analysis is sufficient in most cases, but can rapidly become precarious if the chasm widens between the action plan and the behavior of the machine . . .

Similarly, Sarter, Woods, and Billings (1997, p.1929) reported that:

Empirical research [...] has shown that operators sometimes have gaps and misconceptions in their model of a system. Sometimes operators possess adequate knowledge about a system in the sense of being able to recite facts, but they are unable to apply the knowledge successfully in an actual task context. This is called the problem of “inert” knowledge. One way to eliminate this problem is through training that conditionalizes knowledge to the contexts in which it is utilized . . .

Sarter et al. (1997, p.1929) also added:

When [pilot] responses to this question [i.e. “Are there modes and features of the Flight Management System that you still don’t understand?”] are compared with behavioral data in a subsequent simulator study, there is some indication that these “glass cockpit” pilots were overconfident and miscalibrated about how well they understood about how well they understood the Flight Management System.
Thus, whether knowledge is lacking, is inappropriate, or is inert, the lack of correct conceptualizations and knowledge structures can have significant negative outcomes. Studying the development, measurement, and maintenance of appropriate knowledge and knowledge structures among operators of advanced automated systems is therefore an important issue for the improvement of training operators for complex automated systems.

(4) *Problems related to large resource requirements in working memory.* Working memory stores information along various codes. These include phonetic (verbal), visual (spatial), and sensorimotoric codes. Automated systems rely heavily on working memory as operators must contend with large amounts of variably coded information. Pilots, for example, must simultaneously process phonetic instructions from ATC, build spatial solutions to navigational problems (e.g. by retrieving spatial information regarding the flight profiles in the FMS), and must remember previously obtained information, such as flight briefs and airport restrictions.

While all of the aforementioned working memory codes are necessary for ATC operations (Baddeley & Hitch, 1974), of most importance is how multiple and similar sources of information are coded. It is in this area, in which the resources associated with working memory are particularly limited (Miller, 1956). Interference between memory codes is greater when two codes are presented in the same format (e.g. drawing on the same resources) (Klapp & Netick, 1988). Results such as this, coupled with the work of Wickens’ and colleagues (1983) in stimulus response compatibility for memory codes show the importance of designing systems that minimize interference among the encoding and retrieval of information.

(5) *Problems related to decision making, specifically to the selection of actions to achieve operational tasks.* McClumpha and Rudisill (1994, p. 290) make the point that pilots may often think that they understand the automated systems and their interaction in cockpit tasks with them, but that:

... pilots reported difficulty in knowing what the automation was doing, what the limits and boundaries of performance were, and how to intervene effectively if problems arose. Pilots also reported that many of these types of problems emerge only after considerable line flying or with specific experience in unusual situations. Exposure to these types of situations forces pilots to appraise their ability to understand and feel competent with automation.

Similarly, Woods and Sarter (1993, p. 109) have pointed out:

New capabilities are added, and new options for carrying out tasks under varying circumstances are provided. The ostensible benefit of these additions and modifications of the system is its increased flexibility, which seems desirable in the context of highly complex dynamic environments. However, this flexibility is a mixed blessing as it
often involves immense “cognitive” costs. It imposes new interface management
demands on the user who has to learn more about possible ways of interacting with
the system. He has to understand the benefits and disadvantages of various available
options for a given task, and he has to make decisions as to the relative appropriate-
ness of these options in a particular situation.

We would posit that it is this type of decision making related to the appropri-
ateness of options in a particular situation that is particularly difficult for
pilots. Incidents and accidents show that pilots often fail to adequately assess
the “costs and benefits” of either using the automation or switching it off. As
a results, pilots frequently try to use the automation when they should not,
leading to potentially dangerous outcomes (Amalberti, 1999, p. 178):

Familiarity with onboard computers often prompts people to assume that they can and
should always use them as intermediaries. Pilots tend to get involved in complex
changes to programmed parameters in situations where manual override would be the
best way to insure flight safety. Fiddling with the flight management system (FMS)
makes pilots lose their awareness of the passing of time and, further, their awareness
of the situation and of the flight path (Endsley, 1996; Sarter & Woods, 1991).

Consequently, we believe that knowledge and skill regarding decision
making should be a particular focus of training. In the following section,
we propose some training approaches for this and the other issues
mentioned above.

RECOMMENDATIONS FOR TRAINING FROM THE IP
ANALYSIS OF PROBLEMS IN HUMAN-AUTOMATED
MACHINE SYSTEMS

As Sarter, Woods, and Billings (1997) have pointed out, “. . . training cannot
and should not be a fix for bad design” (p. 1933). Unfortunately, however, two
problems notoriously plague the training issue in automated aircraft. First,
training is often not given adequate attention when new features are introduced.
Edwards noted it already in 1976 (p. 22):

. . . Adequate programmes [sic] of training should be available to introduce new concepts
in control technology. Such criteria are not always given adequate attention prior to the
introduction of yet more gadgetry upon the flightdeck.

Second, even when training is considered, it often is supposed to fill precisely
the role of a fix-everything that Sarter et al. (1997) lament. Amalberti and
Wilbaux, for example, characterized designers’ responses to the well-docu-
mented challenges to the crew aspects of interactions with advanced, automated
aircraft as follows (Amalberti & Wilbaux, 1994, p. 313):
What emerges from these various difficulties is that many situations of poor coordination in a glass cockpit can be related to system design although they are not directly related to a specific action on the interface. This level of causality challenges the philosophy of the system and calls for complex corrections. It is easy to understand that designers are very reluctant to consider that these errors are related to system design and prefer to pass on the problem to trainers.

Given that many problems may be design-related, but that the design is fixed for legacy systems (leaving training as the only reasonable “fix”), the following section is investigating what can be learned from the application of the IP model for training. Specifically, it will be studied to what degree the five major problem areas heretofore identified lend themselves and are ameliorable through training.

(1) **Mismatches Between Displays and the Human Sensory/Perceptual System.** Many of the problems we have discussed relating to two complex systems (automated flight decks and the STARS program) can be accounted for by mismatches between the limitations of the operators’ sensory/perceptual system and the system displays. While the scientific field has done a thorough job in determining the limitations of the human sensory and perceptual system in isolation, what is important here is to determine how the problem is magnified when the operator is performing tasks within the environments with multiple signals and displays. If one was to assume the design of a system is fixed, how might training be advocated to resolve or mitigate these aforementioned problems.

Important system cues must be perceived by the operator. These cues are often relayed within a larger, information-dense, and noisy environment. But it is not the cues in isolation which are difficult to recognize; instead, when presented in competition with other cues, selection of the “right ones” becomes the problem. Consequently, training should emphasize the difficulty in cue competition and train operators to recognize various patterns of cues which equate to determining system status.

Several studies have examined how cue-based training can increase situation awareness. As a summary of this literature, the early work of Mann and Decker (1984) and Bransford, Franks, Vye, and Sherwood (1989) argued that the importance of cueing is based on making the relevant cues distinct from the remaining cues in the environment. Summarizing this work, Stout, Cannon-Bowers, and Salas (1997) reported that cueing distinctiveness can be accomplished in several ways including: (a) displaying the cued behavior in a non-normal context; (b) exaggerating the behavior; (c) continuously repeating the cued behavior; and (d) using the cueing process itself as an opportunity to highlight the behavior. To further illustrate this point, both the theories of naturalistic decision-making
(i.e. recognition primed decision-making, Klein, 1989) and situation awareness (Endsley, 1995) are highly dependent on effective perception of the relevant cues within the environment.

Stout et al. (1997) also discussed four methods of implementation of cue-based training. These included:

1. Passive system prompting – cueing which occurs through passive demonstration
2. Active system prompting – cueing which occurs in real-time
3. Behavioral coaching – passive demonstration cueing accompanied by expert verbalization of attended cues
4. Instructor-guided practice – real-time cueing accompanied by instructor comments of important cues

In summary, we recommend that perceptual training should emphasize the environment in which cues occur, their relative nature with other cues, and the importance of recognizing important cues from non-important cues. This emphasis on cue distinctiveness is at the heart of the Bransford et al. work. The greater the contrast between cues the less attentional resources which must be expended.

(2) *Increased Requirements for Vigilance, Attention-Sharing, and Distribution.* Another problem of automated systems is that the operator’s role is changed from that of an active participant to a passive monitor (Chambers & Nagel, 1985; Hilburn, Molloy, Wong & Parasuraman, 1993; Wiener, 1988). Systems and tasks were originally automated to result in increased accuracy and safety while decreasing workload and costs. This was completed without the foresight of possible downsides of these systems changes. The literature has a long history of human performance decrements during extended vigils (e.g. Mackworth, 1950). In fact, the role of the operator has changed to that of one who must understand system status of all components, have the ability to predict future system state, and understand how the various systems might interact at any moment. Very rarely is the operator of an automated system completely a passive or active participant. Most often the role of the operator is some combination of both, further complicating the multiple tasks at hand.

Training for such increased requirements as vigilance and attention sharing is complex and difficult at best. Work by Parasuraman (Molloy & Parasuraman, 1996; Parasuraman, Molloy & Singh, 1993) has shown the difficulty of human monitoring for failures both single and multiple tasks. Similar findings are present in several ASRS incident (Mosier, Skitka & Korte, 1994) supplying further evidence of monitoring failures for multiple tasks.
Training for vigilance and attention sharing can take several forms. Again holding the notion of system design as an unavailable option (e.g. integrated displays), one interesting finding related to training might be a viable option to relieve vigilance decrements in regard to the monitoring of automated systems. Following the idea of adaptive automation proposed by Rouse (1988), Parasuraman et al. (1993) tested how an adaptive system for functional allocation of automation would perform under various models for a monitoring task. The first model was a non-adaptive group (control group) which remained under automated task control throughout the session. The remaining two groups were both adaptive automation groups. The model-based group received a manual block of automation in the middle of the testing session. For the performance-based group, the middle session was only switched to manual control if a certain criterion was not achieved for the previous sessions. If this performance criterion was not met the monitoring sessions continued under automated control. For all groups, the detection rate for the monitoring task was < 40% before the allocation phase. Following the adaptive allocation phase, the model-based and performance-based group’s performance on the monitoring task increased 50%. There was a slight decrease in the monitoring performance of the non-adaptive control group. These results were replicated for experienced pilots as well.

The research cited in the preceding paragraph illustrates one example of how effective training can counter the problems associated with extended vigils and attention sharing. Disruptions of long periods of automated activity can lead to performance increases by interventions of short periods of manual activity. Not only can this intermittent manual task reallocation improve monitoring performance but it also helps to counteract any degradation of manual skills. Loss of manual skills has been listed as a possible negative consequence of increased automation use (Billings, 1997; Wiener & Nagel, 1988).

(3) Development of Accurate Knowledge in Long-Term Memory. In this report, we have cited numerous examples of poor development of knowledge bases and incomplete mental models. The importance of mental models can be seen in the following definition:

the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states (Rouse & Morris, 1986).

The main point of this definition is the importance of understanding both the current system state as well as having a reasonable ability to predict
future system states. This understanding of a system comes from the current knowledge base of the operator as well as past experiences. A proper understanding of the system at various knowledge levels (e.g. declarative, procedural, and strategic) is important. Users can be trained on either the complete mental model of a system or receive training in some type of sequencing with each model presented building on the last model learned. While research has shown neither method is a guaranteed blueprint for success, the findings do present a distinct advantage for the training of mental models.

Logical reasoning would impart one to believe that to best train a mental model one should expose and teach the components of the target mental model as a whole. Past work in the area of expert/novice differences (see Zsambock & Klein, 1997) would advocate studying the mental models of expert operators for the task under consideration to create a target mental model and then to train its components. In this way, the user is directly exposed to the knowledge of the expert. This interaction usually occurs through some computerized program or database. This approach has been used to train the mental models of geography (Carbonell & Collins, 1973), tactical information about military programs (Crawford & Hollan, 1983), and medical diagnosis (Clancey, 1986). While each of these programs operated in a different manner (self-paced vs. directed learning), they all used the process of presenting the trainee with a complete expert mental model.

Another method for training mental models is to present the user (learner) with a series of mental models. Each model increases in complexity and builds on the knowledge learned from the previous models. Once it is determined the learner has mastered the concepts associated with the current mental model they are exposed to a more complex model. The transitions between each model are meant to be smooth and efficient in reducing errors. Each model is a closer approximation of the final expert model (Campbell, 1990). Frederiksen and White (1987) illustrated the usefulness of this approach by developing a system called QUEST which allowed students to learn a system, troubleshoot problems and work examples for an electronic troubleshooting task. The system allows students to learn the expert mental model through interaction with the system itself.

One training approach which might lend itself as a possible method to effectively train users of the flight management system is called the “training wheels” methodology. This methodology is somewhat different than traditional techniques in that the system is altered to facilitate training and not the instructional materials. Behavioral observation has shown that
new users spent a large amount of time correcting errors instead of learning system components. Consequently, it can be hypothesized that limiting the ability to make these common errors would allow for the users to develop a better understanding of system functions and overall system comprehension. If a certain function is not considered to be essential in the overall operation of basic system tasks and has been found to be a source of common errors for user, this function would be “blocked out” of the training interface, thereby rendering it impossible to activate. This would decrease the time spent in error recovery as the error could not possibly be made in the first place. A recent example of this can be seen in various Microsoft menu structures which “grey out” a function and not allow it to be chosen.

In summary, to create a training wheels interface, one must determine what the essential functions and commands which are needed to perform the primary tasks. Once this step has been determined, the training interface simply disables these tasks, not allowing their activation by the user. If during training, the user attempts to use one of the disabled commands, the system provides a message stating that particular command is not available at this time. All other aspects of the system remain the same. The trainee then uses this system during a training period before being switched to a fully functional system.

(4) Large Resource Requirements in Working Memory. A large proportion of the human-computer interaction errors identified in this report resulted from placing excessive requirements on the limited resources of working memory. Having previously identified these problems, we will now discuss possible training recommendations to alleviate these effects.

One major problem with working memory is due to forgetting, often due to interference of one task with one that has to be remembered. In this context, Wickens and Flach (1988) suggested four steps which can be taken to reduce forgetting due to interference.

(1) Distribute the material to be held in memory over time
(2) Reduce similarity between items
(3) Eliminate unnecessary redundancy
(4) Minimize within-code interference

Each of these suggestions has been empirically derived from extensive testing of the multiple-resource theory. Goettl (1985), for example, determined that acoustically coded, verbal information will have a greater interference with other verbally coded information and spatially coded information will be interfered to a greater extent by spatial rather than other activities.
Yntema (1963), using a dynamic memory task similar to the tasks of an air traffic controller, provided three possible training solutions for overcoming the common limitations of dynamic memory capacity. Following the notion of conceptual chunking, Yntema stated operators are better at recalling a few attributes about numerous objects rather than recalling many attributes of few objects. His second solution stated the amount of detailed information relating to each object poses little threat to the dynamic memory integrity. His final solution stated dynamic memory capacity is enhanced the more distinct the attributes (notice the similarity between this last item and item 2 from the Wickens and Flack listing above). Another possible training recommendation might be found from the work of Egan and Schwartz (1979) on conceptual chunking. They state that if an operator can tie certain characteristics of an object to some concepts in long-term memory, this will increase the capacity of working memory.

(5) Decision Making, Specifically Regarding the Selection of Actions to Achieve Operational Tasks. Training decision-making can be accomplished using several techniques. These include training people to overcome automation biases (Mosier & Skitka, 1996), improve metacognitive strategies (Cohen, Freeman & Thompson, 1997; Jentsch, 1997), and to eliminate the decision-making process through automaticity (e.g. Schneider, 1985). While each of these strategies use a different approach to improving decision-making, it should be noted none of these techniques overcome the limitations of the information processing previously discussed (e.g. limited capacity of working memory).

Mosier and Skitka (1996) have defined automation bias as “the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing”. When an operator suffers from this type of bias it can often lead to errors of omission or commission. Training which makes the operator aware of such a bias specific to automated systems can aid in reducing its occurrence. Other researchers claim operators can become better decision makers if training is received to improve their metacognitive skills. Cohen, Freeman and Thompson (1997) stated that such training would enhance an operator’s ability to: consider appropriate cue for better situation assessment, check situation assessments for completeness and viability for known cues, examine in detail any data which conflicts with situation assessment, and recognize when too much conflict exists between the known cues and the assessment. The researchers tested such a hypothesis with active duty Army officers finding those receiving training tended to consider more factors in their decision process along with a an increased
value of factors illustrating both a quantitative and qualitative increase in the decision-making skills of the operators.

Another training recommendation based more at the raw data level of processing would be to train the relevant cues to the level so the associations between cues and appropriate decisions are automatic (Schneider, 1985). This method is limited in that the cues must always map directly onto one decision and research (Fisk, Ackerman & Schneider, 1987; Schneider, 1985) has shown it takes hundreds of repetitive trials for cue associations to reach an acceptable level of automaticity.

**SUMMARY**

The purpose of this chapter is to provide a first step towards determining the piloting skills needed for the operation of advanced automated aircraft. Using an information-processing (IP) approach, categorization of the problems observed in interactions between human operators and advanced automated systems were determined. Based on these issues, potential training areas were supplied which may aid to ameliorate many of these interactions between humans and automated systems. Following the identification of potential training areas were brief discussions of the potential efficacy of traditional approaches to performance improvements in these areas.

Using the IP framework, numerous problems were identified through examining current literature and research findings, technical reports, as well as incident and accident reports and databases. Five problem areas were identified which, not surprisingly, paralleled the major dimensions of the IP model. These five problem areas included mismatches between hardware and software and the capabilities of the human sensory/perceptual system, increased requirements for vigilance, attention-sharing, and distribution, the development of accurate knowledge in long-term memory, overload of large resource requirements in working memory, and problems relating to decision making. The breakdown of these automation-related problems is important as this step must be completed to determine the foci of any training programs to address these issues. While some believe automation has introduced a whole new array of interaction problems, analyses such as this supply a different snapshot, in our case by using a well investigated information processing model to categorize such problems at their most core level.

Investigations should be conducted to determine the effectiveness of the training approaches we recommended. While Wickens’ IP model is closed-loop in nature, the model does describe a certain path which information follows
throughout the process. Future studies should therefore also examine how training solutions targeting at the earlier stages (i.e. sensation and perceptual) of the model might have a cascading effect in relation to the elimination of problems further along in the processing model.

NOTES

1. Other information-processing models could have been used, but were deemed either too complex to be useful in an airline context (e.g. Rasmussen’s, 1986, IP model) or too simple to be diagnostic (e.g. model).

2. Although working memory is presented here as a single capacity-limited system, its architecture has been conceptualized in more detail. According to Baddeley (1992), for example, working memory consists of a central executive system, to which two subsystems, the articulatory loop and the visuo-spatial sketchpad, are slaved. Each subsystem has its own capacity limits which makes them initially somewhat independent. However, when any subsystem reaches or exceeds its capacity limit, capacity from the central executive is diverted.

3. A “speed bug” is a small mechanical or electronic pointer adjacent to the scale of the airspeed indicator that can be set to indicate targeted airspeed for various aircraft maneuvers and configurations. The “salmon bug” is one of these which indicates the targeted speed during approach in the final approach configuration (i.e. with gear down and flaps/slats extended as appropriate).

4. A flight director gives the pilot steering commands to achieve the desirable flight regime. As a result, the cognitively challenging task of mentally integrating attitude, airspeed, and altitude information is replaced by a somewhat simpler two-dimensional tracking task (i.e. moving the controls so that the airplane vector matches the flight director steering commands).

REFERENCES


7. HUMAN FACTORS ISSUES REGARDING AUTOMATED FLIGHT SYSTEMS

Michael Zuschlag

PROBLEM STATEMENT

Increasing use of automation on air transport flight decks has freed the pilot from much of the work associated with aviating, navigating, and fuel management. However, this does not appear to have reduced the number of accidents or incidents attributable to human error. Instead of removing the operator from the equation, automation in aircraft as in other domains has recast the operator as a monitor and manager of systems, bringing forth the potential for new classes of error (Degani, Mitchell & Chappell, 1998; Federal Aviation Administration (FAA), 1996; Palmer, Hutchins, Ritter & van Cleemput, 1993; Sarter & Woods, 1995a, b; Vakil, Hansman, Midriff & Vaneck, 1995). By studying the characteristics of automated flight systems (AFSs) and the literature associated with error in its use, one can appreciate the sources of these errors and begin to explore means to mitigate it.
DESCRIPTION OF AUTOMATED FLIGHT SYSTEMS

Purpose of Automated Flight Systems

At the most complete level of automation, the AFS flies the aircraft while also performing navigation and fuel management tasks. The AFS flies a complete preprogrammed flight plan, automatically executing course, altitude, and speed changes from point of departure to destination. Most modern AFSs are capable of landing the aircraft. Pilots may use the AFS for both “tactical” and “strategic” control of the aircraft. In tactical control, aircraft behavior is affected on a short time horizon (immediate to several seconds) for lower-order short-term goals (e.g. maintain a given heading). In strategic control, aircraft behavior is affected on a longer time horizon, up to multiple hours, and the goals are more long term and higher ordered (e.g. land on a specific runway following a specific approach procedure).

A flight plan is programmed as a list of waypoints where, at each waypoint, the aircraft is programmed to perform some change in behavior (course, altitude, or speed). The flight plan may include a set of waypoints constituting a standard instrument departure, an arrival, or an approach (precision or non-precision). To execute the flight plan, the AFS automatically determines the aircraft geographic position using a number of sensors (radio navigation, GPS, and/or inertial systems). The AFS then enters a chain of operating modes where each mode flies the aircraft through one element of the flight plan (e.g. a climb, level flight, a turn, a descent).

In principle, with a flight plan programmed, a pilot need only press a button after takeoff and the aircraft will fly the entire route, relegating the pilot to monitoring the system. In practice, weather and air traffic control (ATC) considerations require modification of the flight plan after the aircraft is airborne. Thus, the pilot must maintain constant awareness of the location of the aircraft and the state of the automation. When a change in flight plan is called for, the AFS must allow the pilot to edit the flight plan while in the air or switch to a lower level of automation in which the pilot sets the target flight behavior as each waypoint is passed.

AFSs have the potential to enhance safety by providing accurate navigation and minimizing deviations from a correctly programmed flight plan. Many AFSs also provide “envelope protection” where the AFS senses dangerous states in the aircraft, such as an impending stall, and automatically takes action to prevent a problem (e.g. increase thrust).

Safety was not the only driving force behind the development of AFSs however. AFSs were developed with the intention of reducing cockpit workload
to the point that all functions can be adequately performed by a flightdeck crew of two. Another significant goal of developing AFSs was the desire to capitalize on the capacity of the computer to calculate fuel-efficient cruise altitudes, speeds, and descents.

Components of AFS

The basic components are shown in Table 1 (see also Fig. 1). Different manufacturers use different names for roughly the same components. In this paper, Boeing terminology and abbreviations are arbitrarily selected to refer the generic components.

With these components, the pilot may chose one of five levels of automation for flying the aircraft:

1. **Manual.** Aircraft is hand flown entirely with minimal automation (in the case of fly-by-wire aircraft, there is always some automation). Pilot maintains a desired direction by controlling aircraft’s attitude and thrust through the control column and thrust levers respectively.

2. **Flight director (FD).** Aircraft is hand flown as in manual, but the automation uses the FD to indicate the attitude the pilot should direct the aircraft towards.

3. **Control wheel steering.** Autopilot maintains the attitude last held by the pilot. Pilot is free from making moment-to-moment corrections to maintain an attitude, but can adjust the autopilot to seek to a new attitude.

4. **Autopilot.** Autopilot seeks the attitude and thrust necessary to maintain the flight parameters entered in the mode control panel (MCP) by the pilot. These include altitude, climb/descent rate or angle, heading or track, and speed. The autopilot maintains these flight parameters until the pilot changes them.

5. **Flight Management System (FMS).** FMS flies the aircraft from waypoint to waypoint, making necessary changes to altitude, track, and speed as programmed in the flight plan. The system typically automatically reverts to the autopilot level of automation if the FMS comes to a discontinuity or to the end in the flight plan.

The Dominant Role of the FMS

The FMS is a collection of computers and display-control units that comprise most AFS functions. It contains the flight plan, which may be either entered by the pilot or retrieved from nonvolatile memory. It includes a database of airports, navaids, and other waypoints used to create flight plans and navigate the aircraft. It automatically selects navigation sensors to use and specific
Table 1. Basic Components of an AFS.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbrev</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Systems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autopilot</td>
<td>A/P</td>
<td>Automatically maintains a pilot-entered vertical or horizontal direction.</td>
</tr>
<tr>
<td>Autothrottle</td>
<td>A/T</td>
<td>Automatically controls thrust to maintain a given forward or vertical speed.</td>
</tr>
<tr>
<td>Flight Management Computer</td>
<td>FMC</td>
<td>Controls the aircraft’s flight among a set of waypoints. On an Airbus the FMC, A/T, and A/P are all a single system known as the Flight Management Guidance Computer (FMGC), but the FMC function is still treated separately.</td>
</tr>
<tr>
<td><strong>Predominantly Controls</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode Control Panel</td>
<td>MCP</td>
<td>A control panel on the glare shield that controls the autopilot and takes other tactical input. Called a Flight Control Unit (FCU) on Airbus aircraft.</td>
</tr>
<tr>
<td>Control Display Unit</td>
<td>CDU</td>
<td>A keypad and display panel located on the center pedestal. It controls the FMC where it may be used for either tactical or strategic input. Called a Multifunction Control Display Unit (MCDU) on Airbuses and Boeing 747s where it may perform functions in addition to controlling the FMC.</td>
</tr>
<tr>
<td><strong>Predominantly Displays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Flight Display</td>
<td>PFD</td>
<td>Centrally located graphical symbolic display indicating the aircraft state including attitude, speed, altitude, flight director commands and engaged and armed AFS modes for flight control.</td>
</tr>
<tr>
<td>Flight Director</td>
<td>FD</td>
<td>A command display on the PFD indicating to the pilot where the AFS is directing the aircraft to go.</td>
</tr>
<tr>
<td>Navigation Display</td>
<td>ND</td>
<td>Centrally located electronic moving map display, showing flight plan and actual course. Sometimes called an Enhanced Horizontal Situation Indicator. The ND and PFD combined constitute the Electronic Flight Instrument System (EFIS). Some aircraft are retrofitted with an AFS lack a ND.</td>
</tr>
<tr>
<td><strong>Combined Systems, Displays, and Controls</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flight Management System</td>
<td>FMS</td>
<td>The CDUs and FMCs combined. Also called a Flight Management Computer System (FMCS) on some Boeing aircraft.</td>
</tr>
</tbody>
</table>

nav aids to tune to, and calculates the aircraft’s geographic position from the sensors, sending commands, directly or indirectly, to the control surfaces and throttles to maintain or alter aircraft direction. By default, it uses an altitude, top of descent, and speed that optimizes a combination of fuel economy and schedule keeping collectively indicated by a Cost Index (CI) value entered by the pilot. It keeps track of fuel use and fuel remaining, and can alert the pilot
of such conditions as when the pilot should leave a holding pattern if she or he is to have adequate fuel reserves (Buffer, 1999; Smith Industries, 1996). In some sense, the FMS can be thought of as an autopilot for the autopilot, telling it when to change course, speed, and altitude, although, in the actual architecture, the FMS is not necessarily controlling the autopilot per se.

The CDUs are the control and display components of the FMS (see Fig. 2), although the FMS is also affected by control inputs to the MCP and elsewhere (typically to override settings in the CDU). Air transports typically have three CDUs. One is not part of the FMS but instead is used for other things requiring alphanumeric input such as the Aircraft Communications Addressing and Reporting System (ACARS). The remaining two CDUs, typically located on each side of the thrust levers, normally control a pair of FMCs where an input on either CDU goes to both FMCs. In normal operation, one FMC is algorithmically slaved to the other and the two FMCs check each other to provide reliability through redundancy. The system may allow the pilots to switch the CDUs to operate each FMC independently or to both operate a single FMC in the event one FMC fails.

Fig. 1. Flightdeck of a 747–400 showing some AFS Components. (André Holland)
The CDU has a character-only display that is between $10 \times 20$ to $14 \times 24$ characters in size. CDUs have a combined form and menu user interface\(^2\). Users select a form field or menu item with soft keys ("Line Select Keys" or LSK) along the sides of the CDUs display. This is actually a faster method than can be provided with cursor keys, but parallax considerations limit the display to only ten to twelve selectable items on the screen at once\(^3\). On the CDUs of air transports, the method of entry is unusual for form interfaces: a user first enters the value onto a "Scratch Pad" (SP) line at the bottom of the CDU display and \textit{then} selects the field to enter it into with an LSK.

\textit{Fig. 2.} CDU from a Boeing 757. (Zuschlag)
In addition to an alphanumeric keypad, the CDU has specialized keys, usually located above the alphanumeric keypad. Often called “mode keys,” these set the interface mode of the CDU only; they do not engage any operating modes of the FMS. By pressing one of the interface mode keys, the user accesses a collection of form pages on the CDU. Each key, and thus each page collection, corresponds to a functional area of the FMS. The exact division of functions among these keys varies considerably from aircraft to aircraft, but common important functional areas are as follows (Avsoft Systems, 1989; Buffer, 1999; Smith Industries, 1996):

(1) **Initiation.** Allows a pilot to set aircraft data prior to departure. The pages for this may come up automatically when CDU is turned on.

(2) **Flight Plan.** Allows a pilot to view and edit the flight plan. On Boeing aircraft this function is divided into the “Legs,” “Route,” and “Dep/App” page collections while Airbus FMSs have an “F-Pln” page collection.

(3) **Progress.** Displays current state of aircraft while enroute. Often labeled “Nav” on general aviation (GA) FMSs such as the Universal UNS-1B (ACBI, 1995).

(4) **Phase.** Allows a pilot to view and edit parameters for a particular phase of flight (being climb, cruise, descent, and approach). May be labeled “VNav” or “Perf,” or there may be separate keys for each phase.

(5) **Direct To.** Allows a pilot to command the aircraft to fly directly to a specific waypoint, skipping any other waypoints in between. May be integrated with the flight plan pages.

(6) **Hold.** Allows a pilot to define, arm, and execute a holding pattern. May be accessed from the flight plan page collections in some aircraft.

(7) **Fuel.** Allows a pilot to define and monitor calculated fuel levels and use.

(8) **Data.** Allows a pilot to view details of any waypoint, navaid, or airport. Sometimes labeled “Ref” or “Fix.”

In addition to these keys, a CDU typically has other function keys that generically control the display (e.g. “Next Page” within a page collection). With these page collections, the CDU (and thus the FMS) may be used for either strategic inputs, such as entering a flight plan before departure, or tactical inputs, such as performing a Direct To or Hold at the current position.

**Functional Hierarchy**

An inventory of typical AFS functions is given below in the form of a hierarchy where further-in items are subfunctions of further-out items (ACBI, 1995; Avsoft Systems, 1989; Buffer, 1999; Smith Industries, 1996). Individual
AFSs may have additional or alternative functions. Some functions correspond to operating modes. Function names in italics below indicate operating modes that are typically annunciated to the pilot, usually via the PFD. An AFS also has many other functions that do not directly correspond to modes whether annunciated or not.

(1) **Maintain entered direction.** These are functions traditionally associated with the autopilot. They are typically controlled through the Mode Control Panel on the glare shield.

1.1 **Heading.** Maintain heading or track entered into the MCP (some aircraft allow the pilot to choose heading or track). Sometimes called “Heading Select” Mode. Typically abbreviated as “HDG” on the PFD.

1.2 **Altitude Hold.** Hold at current altitude or an altitude set in the MCP. Sometimes simply called “Altitude”. Abbreviated “ALT” on the PFD.

1.3 **Altitude Change.** These functions put the aircraft into a continuous climb or descent.
   1.3.1 **Flight Path Angle.** Climb or descend at an angle entered into the FPA-V/S control of the MCP. A switch indicates the value is an angle. Abbreviated “FPA.”
   1.3.2 **Vertical Speed.** Climb or descend at the vertical speed (in feet/min) entered in the FPA-V/S control of the MCP. A switch indicates value is vertical speed. Abbreviated “V/S.”
   1.3.3 **Speed.** Climb or descend at an algorithmically determined thrust setting at a vertical rate that maintains a given forward speed. Sometimes called “Open Climb” and “Open Descent.”
   1.3.4 **Expedite.** An Airbus function that commands the aircraft to climb at a maximal possible vertical rate (using maximum continuous thrust) or descend at a maximal allowed speed. Whether for climbing or descending, it is engaged by a single button on the FCU.

1.4 **Autoland.** Land the aircraft automatically. Abbreviated “LAND.”

(2) **Control Speed and Thrust.** These are autothrottle modes.

2.1 **Use maximum thrust for conditions** (*Take off/Go-around*). Abbreviated “TOGA.”

2.2 **Use maximum continuous thrust for conditions** (*Climb*). Abbreviated “CLB.”

2.3 **Use minimum thrust for conditions** (*Idle*).
2.4 **Use a specific value of thrust.** Abbreviated “THR.” Basically, manual mode.

2.5 **Maintain specified speed.** Where speed may be pilot specified on the MCP or CDU, or it may be a constraint in the FMS database, or it may be an FMS calculated optimal speed.

2.5.1 *Speed specified in knots.* Abbreviated “SPD.”

2.5.2 *Speed specified as a Mach number.* Abbreviated “MACH.”

(3) **Provide Envelope Protection.** Prevent the aircraft from entering into dangerous states such as a stall or excessive engine spool down. The degree of envelope protection varies considerably from aircraft to aircraft, with Airbuses, in particular, favoring extensive envelope protection.

(4) **Fly waypoint to waypoint.** These are the functions associated with the FMS.

4.1 **Follow programmed path.** These are the core functions of the FMS. All other functions exist largely to support this function.

4.1.1 **Attitude control while on path.** Essentially, fly the flight plan as programmed.

4.1.1.1 **Lateral Navigation.** Often abbreviated as “LNAV”.

4.1.1.1.1 **Intercept Path and Engage.** When the FMS is armed and the aircraft comes within a certain distance of the programmed flight plan, engage the FMS, intercept the flight plan, and follow it.

4.1.1.1.2 **Track waypoint.** When between waypoints, fly a straight track to the next waypoint. Allowing for the effects of wind, the FMS calculates the heading necessary to achieve the correct track.

4.1.1.1.2.1 **Fly directly on flight plan path.** Fly with no offset.

4.1.1.1.2.2 **Fly offset.** Fly course parallel to entered flight path but to the side by a pilot-entered number of miles. Often used to side-step around weather. See 4.3.9.

4.1.1.1.2.3 **Fly to assume offset.** Fly at an angle off the flight plan in order to begin flying an offset.

4.1.1.1.2.4 **Fly intercept back to flight plan path.** Fly at an angle from an offset back to the flight plan path to cease flying an offset.

4.1.1.1.3 **Turn at waypoint.** When approaching a waypoint with a course change, turn to assume the new course.

4.1.1.1.3.1 **Fly over.** Initiate turn when aircraft is directly over the waypoint. This sort of turn is required for some procedures. See 4.3.8.
4.1.1.3.2 Fly by. Initiate turn shortly before aircraft is directly over the waypoint. This “cutting the corner” makes for a shorter and simpler turn.

4.1.1.4 Fly holding pattern. The holding pattern may be part of a procedure (e.g. to gain altitude), or may be pilot defined either as part of the flight plan or initiated at the present position. See 4.3.7.

4.1.1.2 Vertical Navigation. Abbreviated “VNAV.”

4.1.1.2.1 Profile. Fly vertical trajectory between waypoints. Some aircraft, such as the Boeing 747-400, divide this into the following:

4.1.1.2.1.1 VNav-Path. The elevators maintain a specified vertical profile, analogous to Abbreviated “VNAV PTH.”

4.1.1.2.1.2 VNav-Speed. The elevators control the aircraft speed with no specific vertical path, analogous to. Abbreviated “VNAV SPD.”

4.1.1.2.1.3 VNav-Altitude. Altitude is constrained by the setting on the MCP, overriding any FMS setting. Abbreviated “VNAV ALT.”

4.1.1.2.2 Altitude Change. The Airbus makes a distinction between climb and descending modes. In the Boeing these are unannunciated sub-modes of VNAV-SPD and VNAV-PATH.

4.1.1.2.2.1 Climb. Thrust at a high level and speed maintained with the elevators resulting in a climb. Abbreviated “CLB.”

4.1.1.2.2.2 Descend. Thrust at idle and speed maintained with elevators resulting in a descent. Abbreviated “DES.”

4.1.1.2.2.3 Altitude Capture. Aircraft leveling to smoothly obtain a target altitude. This transition between a vertical movement and level flight may take 10 seconds or more.

4.1.1.3 Fly approach and Standard Terminal Arrival Routes (STARS). These functions involve flying a combination of vertical and lateral directions based on procedures stored in the FMS database.

4.1.1.3.1 Fly ILS (Instrument Landing System) approach.

4.1.1.3.1.1 Capture glideslope and localizer. In localizer capture, the aircraft automatically flies a zigzag pattern hunting for the localizer signal. In practice, it typically finds it on the first zig.
4.1.3.1.2 *Track glideslope and localizer.* Abbreviated “LOC” for the lateral annunciation and “G/S” for the vertical annunciation.

4.1.3.2 Fly *localizer-only approach.* No automated vertical navigation provided.

4.1.3.2.1 Capture localizer. Same as 4.1.3.1.1.

4.1.3.2.2 Track localizer. Same as 4.1.3.1.2.

4.1.3.3 Fly *non-precision approach.* Aircraft flies a flight path angle computed from data stored in the FMS database.

4.1.4 Fly departure. This function involves flying a combination of vertical and lateral directions based on standard instrument departure (SID) procedures stored in the FMS database.

4.1.2 Optimize CI. Cost index is a single number that indicates the relative importance of fuel economy and schedule keeping. Pilots enter the CI when the FMS is initialized for a flight. In principle, the FMS can calculate flight parameters that minimized the overall cost of a flight (in fuel consumption and scheduling delays) based on this number and other data such as winds en route.

4.1.2.1 Calculate best parameters. Calculate optimal speed, altitude and top of descent based on CI entered and other data.

4.1.2.2 Use/Apply best parameters. Apply parameters to flight plan as aircraft is flown. Sometimes said to be the *ECON* mode for the FMS.

4.1.2.2.1 Speed.

4.1.2.2.2 Altitude.

4.1.2.2.3 Top of descent.

4.1.2.3 Override parameters. Allow pilot to override each FMS-calculated parameters with his or her own while still flying with the FMS engaged. Sometimes this can be done individually for each flight phase.

4.1.2.3.1 Speed.

4.1.2.3.2 Altitude.

4.1.2.3.3 Top of descent.

4.1.2.3.3.1 Early descent. Descend before the FMS-calculated top of descent, intercepting FMS-derived vertical profile further down.

4.1.2.3.3.2 Late descent. Descend after the FMS-calculated top of descent, intercepting FMS-derived vertical profile further down.
4.1.3 **Observe constraints.** Fly the flight plan while observing constraints on either speed or altitude.

4.1.3.1 **Hard-coded.** Some constraints are built into the FMS as defaults to allow automatic compliance with various regulations (such as the speed limit below 10,000 feet for U.S. operations).

4.1.3.2 **Database.** Some constraints are stored in the FMS database as part of a procedure such as an approach.

4.1.3.3 **Pilot entered.** Observe any pilot-entered speed and/or altitude constraints for a given waypoint. See 4.3.10.

4.2 **Initialize for Flight.** Prior to takeoff, turn on and input a flight plan and aircraft data to be used during the flight.

4.2.1 **Enter current position and time.**

4.2.2 **Select/enter flight plan.** Air transport FMSs typically allow the pilot to define two flight plans for the same flight. The first is active when LNAV is engaged at takeoff. The secondary flight plan is a back up that may be engaged while en route. It may be used to pre-set an alternative destination or an alternative approach to allow for quick adaptation to changing conditions after takeoff.

4.2.2.1 **Company route.** Air transports typically are delivered with complete flight plans stored in the database for the regularly scheduled routes the aircraft makes. GA FMSs generally do not have this functionality.

4.2.2.2 **Pilot stored route.** High-end GA FMSs typically allow the pilot to store a flight plan she or he has made and retrieve it later to be used again. Some also allow the pilot to reverse the flight plan to fly a return trip. While pilots of air transports can modify a retrieved flight plan, they generally cannot save these modifications for a later flight.

4.2.2.3 **Build from database airways, waypoints, departures, approaches, and airports.** The pilot pieces together or edits a stored flight plan. Changes may be made down to the waypoint by waypoint level. Editing functions and are typically available to aid in this.

4.2.3 **Enter aircraft and fuel parameters.** Allow entry of fuel on board, aircraft weight and center of gravity, wind forecasts along the route and other data used by the FMS in its calculations.

4.3 **Edit flight plan en route.** Allow the pilot to alter the flight plan after airborne to adjust to ATC or weather requirements. Most of these functions can also be done before flight.
4.3.1 **Direct to.** The pilot selects a single waypoint for the aircraft to fly towards. If the waypoint is in the flight plan, the aircraft skips any flight plan waypoints between the present position and the selected waypoint but resumes following the flight plan after the selected waypoint. Sometimes thought of as an operating mode, it is actually more like an editing function in the sense that it practically (and visually on the ND and CDU) deletes all waypoints between the present position and the selected waypoint.

4.3.2 **Change destination.** Allow the pilot to change the destination airport and adjust the flight plan accordingly.

4.3.3 **Activate secondary plan.** Air transports typically allow the pilot to define a secondary flight plan (typically to an alternate airport) before takeoff. This function allows the pilot to activate this plan while en route in the event the primary flight plan is no longer satisfactory.

4.3.4 **Delete waypoint(s) or path segment.**

4.3.5 **Add/Insert waypoint(s) and path segment.** Add a waypoint from the database and insert a connecting flight path segment to the currently active flight plan.

4.3.6 **Create pseudo-waypoint.** Create a temporary waypoint on the flight plan that is not necessarily co-located with any navaid. A pilot may define a pseudo-waypoint by entering a navaid, a radial, and a distance. The FMS treats pseudo-waypoints like any other. A pilot often must do this to get the FMS to comply with ATC instructions to cross a certain radial or DME arc at a certain altitude or speed.

4.3.7 **Holding pattern.**

4.3.7.1 **Define.** Allow the pilot to define a holding pattern for any waypoint or the present position. Pilots enter the waypoint (or “present position”), the outbound radial, the length, and the turning direction (left or right).

4.3.7.2 **Retrieve stored holding pattern.** Select a published holding pattern stored in the FMS database.

4.3.7.3 **Engage.** Begin holding pattern in the case of holding at the present position.

4.3.7.4 **Disengage/ resume flight plan path.**

4.3.8 **Specify Turn Type.** Specify a turn over a waypoint to be either fly-over or fly-by.

4.3.9 **Offset/Crosstrack.** Fly a path parallel to the flight plan path off to the side by a pilot-entered number of miles.
4.3.9.1 **Define.** Pilots enter the number of miles to the side of the flight plan path and the direction of the offset (left or right).

4.3.9.2 **Engage.**

4.3.9.3 **Disengage/resume path.** End flying a crosstrack and re-acquire the flight plan path as described in 4.1.1.1.2.4.

4.3.10 **Altitude and speed constraint.**

4.3.10.1 **Enter.** Enter an altitude or speed constraint for a waypoint. An altitude constraint may be at, at or above, or at or below.

4.3.10.2 **Erase.** Remove a constraint from a specific waypoint.

4.3.10.3 **Override.** This is not necessarily an explicit action, but engaging some operating modes, such as “Expedite,” will cause all constraints to be ignored.

4.4 **Monitor Progress.**

4.4.1 **Display time and distance to waypoint/destination.** A page collection on the CDU that typically shows time, track, and distance to at least the next two waypoints as well as the final destination. May also display an indication of engaged sub-modes such as offset/crosstrack.

4.4.2 **Display fuel use.** Another page collection on the CDU. May also predict fuel use for the remainder of the flight.

4.5 **Navigate.** Calculate geographic position from navigation sensors. FMSs typically combine data from multiple types of sensors in a sort of weighted average. They may also have a list of alternative methods of combining sensor data in the event that a sensor appears to have failed or certain types of navaids are not available.

4.5.1 **Inertial input.**

4.5.2 **GPS input.**

4.5.3 **Radio input.** FMSs typically are capable of area navigation using VORs, DMEs, and other radio nav systems. Algorithms in the FMS search for and select optimal navaids for calculating position (e.g. first look for two DMEs, one ahead and one abeam).

4.5.3.1 **Autotune.** Automatically tune navigation radios to FMS-selected navaids to perform area navigation.

4.5.3.2 **Manual override.** Allow pilot to override automatic navaid selection and specify his or her own selection of navaids.

4.6 **Database retrieval.** Allow pilot to retrieve details about any item in the FMS database (waypoints, navaid, approach, departure, airport info).

4.7 **Alerting.** Display alerting message or symbology concerning unusual events in the FMS or flight plan, such as an inability to maintain required
navigation precision, or an inability to comply with a programmed constraint.

Even a skimming of the above functional hierarchy should indicate the high degree of sophistication found in AFSs, particularly FMSs. While the intention of automation is to reduce workload, the effect is to provide the pilot with an explosion of options for controlling the aircraft, adding to the decision-making workload. Monitoring the FMS is also non-trivial to the degree the pilot must understand its functions in order to understand its behavior.

**Modes**

At the heart of an AFS is its “operating modes.” An operating mode is a way the AFS controls the path of the aircraft. The AFS takes inputs from the environment (e.g. speed, altitude, geographic position, attitude) and, based on this, sends outputs to the aircraft’s control surfaces and other systems in order to reach or maintain some value on the inputs.

An operating mode is to be distinguished from “interface mode,” which is a set of parameters displayed and parameter inputs allowed under certain conditions (e.g. a page collection on the CDU of an FMS). While some operating modes have corresponding interface modes (e.g. HOLD), there is generally not a one-to-one correspondence. Furthermore, even for operating modes that have corresponding interface modes, at any given time the operating mode and the interface mode are independent (e.g. the CDU does not have to be on the VNAV page in order for the aircraft to be flying in the VNAV operating mode).

Formally, at a given instant, the automation state of an aircraft is defined by a set of “base” operating modes (Vakil, Hansman, Midriff & Vaneck, 1995), where each base mode is defined by the following mode elements:

1. A **parameter** to control the aircraft on, e.g. altitude, pitch angle, or thrust level.
2. A **reactor** component, or the specific control surfaces or other systems used to seek and maintain the target value of the parameter, e.g. ailerons, elevators, and thrust.
3. A **source** for the target value of the parameter, e.g. a MCP setting, a field on a page in the CDU, a noneditable field in the database, hardcoded in the FMS.
4. A specific **target** value for the parameter, e.g. 10,000 feet above mean sea level (MSL), 250 knots calibrated airspeed (CAS), 090 degrees track.
For example, suppose the aircraft is in Vertical Speed (V/S) Mode such that the aircraft is climbing at a rate of 2000 feet per minute (fpm). The parameter is Vertical Speed, the reactor may be the elevators (if so designed), the source may be the V/S-FPA setting on the MCP, and the target is +2000 fpm.

Modes differ from one another by one or more of the above mode elements. For example, V/S Mode differs from Flight Path Angle (FPA) Mode in that for V/S the parameter is vertical rate of climb (measured in fpm) while for FPA the parameter is vertical angle of climb (measured in degrees). Two speed modes may differ in that one uses the elevators as the reactor component, while another uses thrust.

The mode elements are listed above in hierarchical order in the sense that a difference between two modes at a higher level on the list represents a more fundamental difference than a difference at lower level on the list. Differences at the lowest level, the target element, are generally not considered different modes at all. For example, whether flying a heading of 090 or 095, the mode is still “Heading Select;” the difference in target is considered just a different setting within the same mode.

Modes may be further defined and distinguished by any of the following:

1. **Conditions to arm.** Events from the environment, the program, or the pilot that set the mode to automatically engage when some other later event occurs.
2. **Conditions to engage.** Events from either the environment, program, or the pilot that will engage the mode.
3. **Normal termination.** Events from the environment, the program, or the pilot that will end the mode and engage another mode in its place as part of routine flying.
4. **Abnormal termination** ("envelope protection"). Boundaries of operation of the mode, which, if crossed, will cause the AFS to automatically disengage the mode and engage another (often intended to be safer) mode. This includes switching to a lower level of automation (or no automation) in some cases. As in any mode change, the new mode is typically announced on the PFD.
5. **Pilot Overrides.** Actions by the pilot that will override some element of the mode or suspend a mode (place it in an armed state) and place the aircraft in a lower level of automation.

At any given time, more than one base operating mode may be active. For example, while the aircraft is climbing in V/S mode, another mode may be active that is keeping forward speed to a specific Mach number, and yet another mode may be controlling the aircraft’s roll and lateral movement.
In addition to base modes, there are “macro” modes (Vakil et al., 1995) which are collections of base modes that are active simultaneously and/or serially. Thus LNAV mode, in which the aircraft flies from one waypoint to another, may be considered to be a macro mode composed of a series of straight tracking modes alternating with turning modes.

While multiple modes may be simultaneously active, there are usually sets of modes that are mutually exclusive. For example, an aircraft cannot be simultaneously climbing at a specified rate of climb and also at a specified flight path angle – the pilot must set the parameter to be one or the other. These sets of mutually exclusive modes constitute a *mode dimension*. At a minimum, an AFS has separate mode dimensions for thrust, lateral movement, and vertical movement. As examples, the annunciated modes for each of these mode dimensions are shown in Table 2 for two aircraft:

Modes may have submodes, where the submode is a varying specification of the mode. The submodes may be different base modes that are used sequentially within a macro mode. Thus, within LNAV mode, there is one submode of tracking straight to a waypoint and another of turning at the waypoint. Submodes may also be modes that differ at lower mode element levels within a superordinate mode that is defined by a high level mode element. For example, within a mode that maintains a constant speed (parameter), two submodes may exist, one for which the source for the target speed is on the Performance Page of the FMS (set for given phase of flight), while for the other, it is on the Legs Page (set for a given waypoint).

A mode may be *inactive* (not currently in effect), *engaged* (currently in effect), or *armed*. An armed mode is not currently in effect but it will be engaged by an event other than pilot intervention that is likely to occur in the current context. For example, when flying with Heading Select engaged and LNAV armed, the aircraft is currently automatically flying the heading entered in the MCP, but when it intercepts the flight plan path, it will engage LNAV and turn and fly the flight plan path. Formally, an armed mode can be defined as a submode of the engaged mode. In this example, Heading Select may be thought of as having two submodes, one with LNAV armed and one with it not.

The transition from one mode to another may be one of three kinds (Vakil et al., 1995):

1. **Commanded.** The mode changes immediately due to an input by the pilot that normally changes the mode in that manner, such as pressing the Altitude Hold button on the MCP to switch to Altitude Mode.
2. **Programmed.** The mode changes due to a plan entered into the system by the pilot beforehand that normally changes the mode in that manner. For
example, the mode changes to Descend as the aircraft approaches an altitude that has a pilot-entered altitude restriction, or the mode becomes LNAV because the pilot armed the mode to engage when the flight plan path is intercepted.

(3) **Uncommanded.** The mode changes automatically independently of pilot programming due to current environmental conditions, including the inability to complete programmed settings.

<table>
<thead>
<tr>
<th>Mode Dimension</th>
<th>A320 Modes</th>
<th>747-400 Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thrust</td>
<td>Take-off/Go-around (TOGA)</td>
<td>Thrust Reference (THR REF)</td>
</tr>
<tr>
<td></td>
<td>Flex take-off (FLX)</td>
<td>Hold</td>
</tr>
<tr>
<td></td>
<td>Max Continuous Thrust (MCT)</td>
<td>Idle</td>
</tr>
<tr>
<td></td>
<td>Climb (CLB)</td>
<td>Thrust (THR)</td>
</tr>
<tr>
<td></td>
<td>Idle</td>
<td>Speed (SPD)</td>
</tr>
<tr>
<td></td>
<td>Thrust (THR)</td>
<td>Speed (SPD)</td>
</tr>
<tr>
<td></td>
<td>Speed (SPD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mach</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alpha Floor (stall protection)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOGA Lock (TOGA LK)</td>
<td></td>
</tr>
<tr>
<td>Lateral Movement</td>
<td>Runway (RWY)</td>
<td>Take-off/Go-around</td>
</tr>
<tr>
<td></td>
<td>FMS controlled (NAV)</td>
<td>Heading Hold (HDG HOLD)</td>
</tr>
<tr>
<td></td>
<td>Heading (HDG)</td>
<td>Heading Select (HDG SEL)</td>
</tr>
<tr>
<td></td>
<td>Track (TRK)</td>
<td>FMS controlled (LNAV)</td>
</tr>
<tr>
<td></td>
<td>Localizer Capture (LOC*)</td>
<td>Localizer (LOC)</td>
</tr>
<tr>
<td></td>
<td>Localizer (LOC)</td>
<td>Roll out</td>
</tr>
<tr>
<td></td>
<td>Approach (APP)</td>
<td>Attitude (ATT)</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roll out</td>
<td></td>
</tr>
<tr>
<td>Vertical Movement</td>
<td>Speed Reference Scale (SRS)</td>
<td>Take-off/Go-around</td>
</tr>
<tr>
<td></td>
<td>Open Descent (OP DES)</td>
<td>Altitude Hold (ALT)</td>
</tr>
<tr>
<td></td>
<td>Open Climb (OP CLB)</td>
<td>Vertical Speed (V/S)</td>
</tr>
<tr>
<td></td>
<td>Expedite Descent (EXP DES)</td>
<td>Flight Level Change – Speed (FLCH SPD)</td>
</tr>
<tr>
<td></td>
<td>Expedite Climb (EXP CLB)</td>
<td>VNav-Path (VNAV PTH)</td>
</tr>
<tr>
<td></td>
<td>Flight Path Angle (FPA)</td>
<td>VNav-Speed (VNAV SPD)</td>
</tr>
<tr>
<td></td>
<td>Vertical Speed (V/S)</td>
<td>VNav-Altitude (VNAV ALT)</td>
</tr>
<tr>
<td></td>
<td>FMS controlled Descent (DES)</td>
<td>Glide Slope (G/S)</td>
</tr>
<tr>
<td></td>
<td>FMS controlled Climb (CLB)</td>
<td>Flare</td>
</tr>
<tr>
<td></td>
<td>Altitude Capture (ALT*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Altitude Hold (ALT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glide Slope (G/S)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td></td>
</tr>
</tbody>
</table>
Under certain conditions, the AFS may automatically ignore portions of the programmed flight plan, substituting its own flight performance instead, based on its own internal algorithms and parameters. For example, in the case of the Airbus 320, the AFS will prevent the aircraft from descending for a waypoint when it is generally climbing, and vice versa. Specifically:

1. If the pilot programs an altitude constraint that requires a descent while the aircraft is in Climb Phase Mode, then the constraint is ignored.

2. If VNAV is engaged after the aircraft has climbed above the altitude constraint for a downpath waypoint, then that altitude constraint is ignored.

3. If the pilot programs an altitude constraint that requires a climb while the aircraft is in Descent Phase Mode, then the constraint is ignored.

4. If VNAV is engaged after the aircraft has descended below the altitude constraint for a downpath waypoint, then that altitude constraint is ignored.

5. If aircraft passes the top of descent point and it is already below the altitude constraint of the first waypoint, it does not attempt to make that constraint but instead flies at half the flight plan’s flight path angle or at –500 fpm (whichever is steeper) until it intercepts the flight plan profile.

6. Speed and altitude constraints are ignored when using Expedite.

The Airbus AFS notifies the pilot of these deviations from the flight plan by placing an orange asterisk beside the constraint that is ignored on the Flight Plan Page in the CDU (which is not necessarily displayed), except for the last case in which the only feedback is the PFD annunciator indicating that Expedite is engaged. The only way for the pilot to prevent these AFS overrides is by going to a lower level of automation, e.g. by entering the constraint altitude in the FCU and disengaging VNAV.

**PROBLEMS ASSOCIATED WITH AFSs**

In a review of incidents and accidents documented in Palmer (1995) and Palmer, Hutchins, Ritter, van Cleemput (1993), flight automation operation has contributed to the cause of incidents and accidents for five basic reasons. These five reasons may be classified into three categories as follows:

1. Operating mode Problems
   a. Mode Awareness
   b. Mode Prediction
(2) Programming Problems
   a. Programming Error
   b. Programming Awkwardness
(3) Diagnosis Difficulty

Operating Mode Problems

These are problems associated with the AFS operating modes, i.e. the automated flight modes that control the path of the aircraft. (For details, see Table 3.)

Attention is recently being directed at mode awareness (Chappell, Crowther & Mitchell, 1997; Degani et al., 1998; Palmer, 1995; Sarter & Woods, 1997; Vakil, Midriff & Hansman, 1996), with research suggesting that just knowing the name of the current mode may not be sufficient to prevent problems. Pilots must also be able to anticipate mode changes, including uncommanded changes that may place the aircraft in a dangerous attitude.

Table 3. Operating mode Problems of AFSs.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Definition</th>
<th>Contribution to Incidents and Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode Awareness</td>
<td>Crew does not notice or correctly perceive signals indicating an uncommanded or programmed change in a mode or mode parameter, or that a mode had failed to engage or change as intended. This includes cases of the crew incorrectly identifying a signal or annunciator from the AFS.</td>
<td>The crew no longer knows the operating mode of the aircraft. Crew inputs may thus have unintended effects sending the aircraft onto an undesirable path. If the crew is unaware of the path, or unable to countermand it, an incident or accident may result.</td>
</tr>
<tr>
<td>Mode Prediction</td>
<td>Crew did not anticipate an uncommanded change or failure to change in the AFS’s mode or parameter for a mode. Crew not knowledgeable of or forgot about this aspect of the AFS.</td>
<td>The aircraft may suddenly enter a mode that the crew is not prepared to deal with. In extreme cases the aircraft may go out of control. Frequently, this problem is combined with mode awareness problems, i.e. the crew did not anticipate a mode change and did not notice it when it did change. In this case, its contribution to an incident or accident is the same as with a mode awareness problem.</td>
</tr>
</tbody>
</table>
Also, pilots must not only know the identity of the mode, but also must know what the mode means – exactly what the aircraft is doing or going to do given the mode’s parameters, reactors, sources, targets, and conditions for transition into other modes. Pilots need to be made aware when a mode fails to do what it is supposed to do, either due to a failure or an uncommanded mode transition. Unfortunately, this may represent too much information for a human to reliably memorize or monitor. Perhaps it would suffice if pilots could be kept aware of aircraft’s current and predicted behavior at a more fundamental level than the operating mode per se.

**Programming Problems**

These are problems associated with the user interface for programming an FMS or other components of the AFS (see Table 4).

While much attention is justifiably directed at operating mode problems, programming problems are at risk of being overlooked. It appears that programming problems are the most frequent contribution to incidents (Vakil et al., 1995). While in principle most programming is done before flight when the pilot can devote full attention to it, changes to the flight plan en route are most likely during high workload periods (Wiener, 1989). Changing an approach while nearing the terminal area is frequently required and notoriously difficult on FMSs.

**Table 4.** Programming Problems of AFSs.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Definition</th>
<th>Contribution to Incidents and Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming Error</td>
<td>The crew makes an entry error into the AFS that is not caught until after the aircraft deviates from the intended path.</td>
<td>Aircraft unknowingly deviates from intended path, where potentially it may collide with terrain or another aircraft.</td>
</tr>
<tr>
<td>Programming Awkwardness</td>
<td>Programming is difficult and slow. This may be because the crew is repeatedly unable to correctly enter an intended flight path (here, making incorrect entries does not result in actual path deviations), or it may be simply because normatively programming the AFS takes many keystrokes.</td>
<td>Programming time is head-down time that prevents one or both pilots from monitoring other things necessary to fly the plane correctly. Programming awkwardness is a problem likely to cause both pilots to go head down as one pilot tries to help the other correctly program the AFS. Pilots may thus not notice traffic or other problems with other systems.</td>
</tr>
</tbody>
</table>
It also appears that the FMS user interfaces could use considerable human factors attention (Hughes, 1997; Sherry, Feary, Polson Mumaw & Palmer, 2001). In an unpublished heuristic review of one FMS, over 100 potential usability issues were identified, and that did not even include all interface modes.

**Diagnosis Difficulty**

**Table 5.** Diagnosis Difficulty with AFSs.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Definition</th>
<th>Contribution to Incidents and Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis Difficulty</td>
<td>Crew trying to diagnose a problem with the AFS, being either that the crew cannot program the AFS as desired or that the AFS has done something undesirable. This distracts the crew from other things necessary to fly the plane correctly.</td>
<td>All diagnosing time is head-down time, which prevents one or both pilots from monitoring other things necessary to fly the plane correctly. Sometimes this causes both pilots to go head down as both try to correctly diagnose the AFS. In extreme cases, the crew becomes totally absorbed in the AFS problem, and maybe oblivious even to warning signals from other systems.</td>
</tr>
</tbody>
</table>

**Summary**

There are a total of five classes of problems associated with AFSs, The five classes can be roughly mapped to Wiener’s (1989) three questions that operators need answered about a piece of automation, as shown in Table 6.

**Table 6.** Relation of Problemss to Wiener’s (1989) Components of Automation Awareness.

<table>
<thead>
<tr>
<th>Component of Automation Awareness</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is it doing now?</td>
<td>Mode Awareness</td>
</tr>
<tr>
<td>Why is doing that?</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>What will it do next?</td>
<td>Mode Prediction</td>
</tr>
</tbody>
</table>

*To expand Wiener’s list to include automation operation as well as monitoring, one may also add:*

| How do I get it to do something else? | Programming Difficulty and Programming Error |
SOURCES IN DESIGN

The above problems in the operation of AFS can be traced to sources within the design of the systems themselves.

Inconsistent Interfaces

The need for higher consistency among various AFSs is well documented (FAA, 1996). A lack of consistency can contribute to any of the above problems, especially as pilots move from one aircraft to another and confuse the operations of each. Comparing among automated AFS-equipped aircraft, there is a lack of consistency in the items listed in Table 7.

Generally, inconsistency is greater across aircraft rather than across avionics manufacturers. This is true even when the same avionics manufacturer makes the equipment for different aircraft. Honeywell, for example, makes the FMSs for both the A320 and the Boeing 777, while these are least like each other. Broadly speaking, the user interfaces for FMSs can be categorized by similarity into three families:

- Boeing (“Brown”, so called because of the FMS chassis color on larger Boeing aircraft).
- Airbus and McDonnell Douglas (“Gray”).
- High-end GA, such as the Allied Signal and Universal FMSs.

All Airbus FMSs are highly similar to each other due to Airbus’s commitment to type rating across aircraft. Within the Boeing fleet, the 767 uses the same FMS as the 757. The 737 FMS (made by Smith) appears very similar to the 757/767 FMS (made by Honeywell). The 777 and 747 are relatively different from them and from each other.

While inconsistency is greatest across aircraft, there are also inconsistencies across FMSs within the same aircraft model owing to hardware and software upgrades. There are many versions of FMS software currently flying on 737s for example, while training only covers one version. In the older FMSs of the Airbus A320, pilot modifications to the flight plan go into effect immediately after being made. In the latest release of the software for the A320, this is true only for the Direct To function; all other modifications do not go into effect until the pilot presses an “Insert” soft key.

There are also inconsistencies within a given FMS of a particular version for a particular aircraft. In the A320, for example, the 6L line select key on the Performance Page activates the approach for some phases while it displays the performance page for the previous flight phase for other phases. If a pilot
Table 7. Inconsistencies in AFS Interfaces.

<table>
<thead>
<tr>
<th>Item</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nomenclature</td>
<td>The glareshield automation panel is known as a Mode Control Panel on Boeing aircraft, while it is a Flight Control Unit on Airbus aircraft.</td>
</tr>
<tr>
<td>Operating Modes</td>
<td>The Boeing 747-400 has the vertical modes VNAV-Path and VNAV-Speed, either of which may be used for climbing or descending. The A320 has separate Climb and Descend vertical modes.</td>
</tr>
<tr>
<td>Symbols</td>
<td>The NDs of the A320 and the Boeing 777 have entirely different symbols for the top of descent.</td>
</tr>
<tr>
<td>Control Mechanics</td>
<td>On the Boeing 777 pushing the Heading/Track knob engages Heading Select mode. On the A320 pulling the Heading/Track knob engages Heading Select; pushing it engages LNAV.</td>
</tr>
<tr>
<td>Symbols</td>
<td>On the A320, the secondary flight plan has white lettering, while the active flight plan has green lettering. On the Boeing 777, the active flight plan has white lettering. Green is used to indicate the current value of multi-selection fields.</td>
</tr>
<tr>
<td>Input Syntax</td>
<td>FMSs for transports use a value-field syntax through use of a “scratch pad”. FMSs for high-end GA (such as the Allied and Universal systems) use field-value syntax where the pilot directly types on a field.</td>
</tr>
<tr>
<td>Application of Soft Keys</td>
<td>On the newest A320s, a flight plan change is activated or discarded using line select keys. In the Boeing 777, activation is done with a dedicated Execute key, while the change is discarded with a line select key.</td>
</tr>
<tr>
<td>FMS Page Collection Divisions</td>
<td>Boeing aircraft have a separate key for approaches and departures. The pilot of an A320 accesses this information through the flight plan pages.</td>
</tr>
<tr>
<td>Progress Display</td>
<td>On the Progress Page of the FMS, the Boeing 747 displays previous waypoint for the current position, the next two waypoints for the current position, and the destination. The Boeing 777 displays the next two waypoints from the current position and the destination but not the previous waypoint. On the A320 Progress Page the pilot may select only one point at a time to display while the previous and next five waypoints are displayed on the F-PLN page.</td>
</tr>
<tr>
<td>Flight Plan Scrolling</td>
<td>Keys for the Boeing FMSs scroll the flight plan down one page at a time. Keys for the Airbus scroll it one waypoint at a time.</td>
</tr>
<tr>
<td>Flight Plan Editing Methods</td>
<td>The A320 has “Revision” pages used to modify flight plans. The Boeing 777 relies on a copy-and-paste functionality of the scratch pad to effect changes.</td>
</tr>
<tr>
<td>Support of Victor and Jet Airways</td>
<td>In the Boeing 757 and 767 Pilots may program a flight plan using airways on the Route pages. The A320 only allows programming by individual waypoints.</td>
</tr>
</tbody>
</table>
backtracks from one performance page to another too quickly, she or he may activate the approach. In another example, pre-selected speeds become the Automatic Speed only for the Descent Phase and not for other flight phases. In the Universal UNS 1B, in order to delete something, sometimes the user presses a “Delete” soft key twice (first to initiate, second to confirm), but other times a single press performs a delete, and a second press undoes the delete.

Efforts towards consistency may address these concerns. Such efforts may also open an opportunity for including other user interface features such as a consistently operated and consistently accessible Undo capability, and the entry of field values in a manner similar with the way clearances are received from ATC (FAA, 1995; Proctor, 1997; Riley, DeMers, Misiak & Schmalz, 1998).

Operating Mode Proliferation

By this paper’s definition, an operating mode is a mapping between inputs to the aircraft and reactions in the form of the path taken by the aircraft. Each mapping is among a set (i.e. mode dimension) of alternative mutually exclusive mappings.

Problems with operating mode awareness and prediction are aggravated by the large number of operating modes on modern automated aircraft (Sarter & Woods 1995a). The more modes possible, the larger the number of possibilities a pilot must consider. The 747-400 has 21 annunciated operating modes. The A320 has 28. These modes are spread across thrust, vertical, and lateral mode dimensions, so up to three modes are simultaneously active. In the A320, there are also modes associated with the dimensions of flight phase (Preflight, Takeoff, Climb, Cruise, Descent, Approach, Go Around, Diversion, and Done) and radio navigation (automatic, bias, and manual).

Many modes have multiple sub-modes. For example, Approach Mode on the A320 may be full ILS, Localizer Only, or Nonprecision. There are also three different Holding modes, varying on the means by which they are discontinued. For the A320, the total number of modes, including all sub-modes, is in the neighborhood of 80, depending on what is considered a “mode.”

A lack of mode awareness may be most frequent at the lower mode elements (i.e. source or target), as this would represent subtle aspects about a mode that are easy to miss. However, when there is a lack of awareness at this level, it is may not be as serious as the less frequent case of lack of awareness at higher levels. For example, in the cruise phase, knowing the aircraft is descending rather than holding an altitude (parameter mode element) is probably more important than knowing the precise rate of descent (target mode element). Indeed, it appears that accidents (but not necessarily incidents) are more often
associated with a lack of awareness of high level mode elements (e.g. mistaking V/S for FPA). In a classification of the accidents listed in FAA (1996), for example, it appears that mode awareness at the parameter or reactor level was implicated in up to 14 out of the 24 accidents.

**System Insulation**

Some problems with operating mode awareness and diagnosis difficulty may be traced to “system insulation,” where the automation makes it more difficult for the pilot to sense what the aircraft is doing. Being less aware of the aircraft’s state is, by definition, a lack of operating mode awareness. Also, the degree the system inhibits the pilot from knowing the aircraft’s state is the degree this system insulation will interfere with diagnosis. At the very least, the pilot will need time to come up to speed on the aircraft’s state before engaging in productive diagnosis.

Some features in current AFS design tend to aggravate system insulation.

- **Uncommanded Mode Changes.** When an AFS automatically switches to a new mode, feedback to the pilot is typically limited to a visual change of the mode annunciators on the PFD. Some aircraft may also provide an audible click (Mecham, 1994). Pilots who fail to check the mode annunciators regularly may execute an action that was appropriate for the former mode but is inappropriate for current mode or fail to commit an action necessary for the mode (Sarter & Woods, 1995b). The Paris Airshow crash of an A320 is one example of a fatal accident associated with the AFS switching automatically into a mode without pilot awareness (Casey, 1993). In effort to deal with this through training, FAA (1996) recommends that pilots be informed of “the conditions under which the [automation] will or will not engage, will disengage, or will revert to another mode” (FAA, 1996 [Automation Mgt-2]).

- **Programmed Mode Changes.** Programmed mode changes can have the same effect as uncommanded mode changes. One might expect programmed mode changes to be less problematic since the pilot had programmed the change and thus should be expecting it. However the lag since programming may be both long (possibly hours) and difficult to predict. One pilot error associated with programmed mode changes is so common that it even has a name: a “kill the capture” altitude deviation (Palmer, 1995). When an aircraft is changing altitude through one of the vertical operating modes, the pilot can typically adjust the rate of vertical movement with the MCP. However,
if this done \textit{while the aircraft is leveling} at the target altitude (in “Altitude Capture” mode), then the aircraft will cancel leveling the aircraft and resume vertical movement taking the aircraft beyond the intended altitude. While, the Altitude Capture mode is annunciated on the PFD, the PFD is not in the foveal field of view when the pilot is reaching for the MCP. Thus “killing the capture” is a relatively common error.

- **Flight Plan Automatic Overrides.** These potentially can cause similarly described problems as uncommanded and programmed mode changes, although they do not appear to show up in incidents very much.

- **Acceleration Minimization and Lack of Vestibular Input.** Most AFSs fly trajectories that minimize vertical accelerations to about 0.05 g in order to maximize passenger comfort. This, however, also removes vestibular feedback to pilots of significant attitude changes associated with mode transitions, such as between climbing and Altitude Capture.

- **Fly-by-wire Without Feedback.** Some fly-by-wire aircraft do not move the control column/side stick or thrust levers when the AFS controls flight. Some pilots rely on the feedback provided by these control motions to monitor the automation (Bluecoat, 1997a; Sellen, Kurtenbach & Buxton, 1992).

- **Multi-mode Displays Obscuring Data.** The ND and particularly the CDU can each be set to display different information from one time to the next. There are over 30 “pages” of information in the A320 or 777 CDU for example, only one of which is visible at a time. This means that a lot of information is literally out of view while flying. A pilot attempting to diagnose a problem with the AFS may have to hunt through several pages of information on the CDU.

- **Failure Masking.** Several incidents and accidents are associated with the AFS masking a failure of the flight system by automatically (and very smoothly) compensating for it (FAA, 1996). In the usual scenario, the automation compensates to the limit of its envelope at which point it disengages suddenly tossing a nearly out of control aircraft into the lap of the pilot.

- **Separate and Unlinked Electronic Controls and Dual Operators.** System insulation can also occur whenever one pilot in the crew is unaware of inputs made by the second pilot to the AFS. This robs the crew of a chance to cross-check. This is most often a serious problem with inputs on the CDU, as the CDU for one pilot is not in easy view for the other pilot (unlike the MCP, which is in view for both pilots). CDUs can be slaved so that input on one is “echoed” on the other so both pilots always see the same thing. However, it is often more desirable for each pilot to display different pages on their respective CDUs to maximize the amount of the system in view and under control at any given time.
• **Complacency.** The very high reliability of the AFS that is necessary for its certification for air transports can engender complacency in the aircrew, encouraging the reallocation of attention and effort from monitoring the AFS to other cockpit tasks (see Small Keyhole section).

Some system insulation is inherent in automation as the entire purpose of automation is to free the operator from moment to moment involvement with the system. An AFS would have little value otherwise. However, at the very least, the system insulation must be low enough to allow the operator to detect failures in the automation or any case when the system does not perform as intended by the operator. There are also two specific conditions when system insulation must be minimized, both involving operator intervention:

1. When the operator changes the functioning of the automation, whether taking over full control or merely making an adjustment.
2. When the operator attempts to diagnose why the system had not performed as intended.

For instance, in regards to the “kill the capture” altitude deviation, it may not be strictly necessary for the pilot to know exactly when the AFS transitions to Altitude Capture mode. Instead, the pilot just needs to know the mode at the time she or he reaches to adjust the vertical speed.

---

**Distributed Controls and Displays**

The controls and displays that affect the AFS are distributed around the cockpit in a manner that may not be optimally organized and may contribute to several of the problems of AFSs, namely reducing mode awareness, increasing programming errors and diagnosis difficulty. In current flight deck design, controls are distributed by the underlying system they are associated with. All FMS control is through the CDU, for example, while all autopilot control is through the MCP. Radio navigation control is on the central pedestal, and the fuel system is on the overhead panel. Perhaps it would be better to organize controls and displays by function (e.g. aviating, navigating, etc.) rather than system.

It has already been mentioned that the mode controls on the MCP are relatively distant from the mode annunciators on the PFD for example. This physical distribution is particularly pronounced for functions available through the CDU and respective functions available through conventional physical controls. For example, while target altitudes may be entered in various places among the pages in the CDU, an altitude entered in the MCP still has effect (typically overriding CDU-enter values) even when the FMS...
is controlling vertical navigation. In the A320, selecting an approach in the CDU when the aircraft nears an airport will not arm the approach unless either “ILS” or “LOC” is armed on the MCP. Radio tuning likewise may be manually set in the CDU or overridden from the radio console (Buffer, 1999).

In some cases the coupling between the CDU and other controls can be quite complex (Sarter & Woods, 1995b). For example, in the A320 among the three suggested ways to automatically fly a localizer-only approach is the following involving three different panels on the flight deck:

1. Select an appropriate VOR approach on the Vertical Revision Page of the CDU.
2. Tune the ILS frequency on the radio console (Radio Management Panel). This overrides the automatic tuning associated with the VOR.
3. Press the LOC button on the MCP. This arms the LOC mode, which the AFS will automatically engage once the localizer beam is captured. (If, instead of LOC, ILS button is pressed on the MCP then a full ILS approach is flown even though a VOR approach was selected in the CDU).

The distribution of controls and displays also occurs within the CDU itself. For example, the target speed for the A320 can be due to any one of the following sources:

1. A default value the FMS follows for certain conditions, such as when the aircraft is below a certain altitude or in a holding pattern.
2. A read-only database value associated with a portion of a particular approach.
3. A calculated value for optimizing the CI based on weather and aircraft characteristics.
4. A calculated speed to prevent breech of envelope protection (e.g. descending too fast or stalling).
5. A “preselected” speed for the flight phase as entered by the user on a Performance Page of the CDU.
6. A pilot-entered speed restriction for a waypoint on the Flight Plan Page of the CDU.
7. A speed entered in the MCP.
8. Whatever speed results from the Expedite Mode.

The actual speed value for each of these conditions, if it is displayed at all, may be on any number of physical controls or CDU pages. Altitude targets are treated similarly.

This distribution of controls and displays potentially contributes to several of the problems associated with AFSs:
(1) **Mode Awareness.** With details of the aircraft performance, such as the actual target value for speed or altitude, being on any number of possible displays, pilots may be looking at the wrong display to infer aircraft behavior.

(2) **Programming Errors.** Pilots may not appreciate the complex interaction of the setting of multiple controls, and thus may set the AFS to do something unintended. The relation between altitudes set in the MCP and the CDU, for example, are sufficiently subtle that one study sought to find a better way to train pilots on it (Chappell, Crowther & Mitchell, 1997).

(3) **Diagnosis Difficulty.** Pilots may have to search through any number of physical and logical places to try to understand why the AFS is not performing as desired. It is all too easy for pilots to focus on just one panel of physical controls or one page of the CDU when diagnosing a problem, and to find it bewildering that the AFS is failing to perform as instructed on that particular display.

*Small Keyhole*

Related to the problems of system insulation and control/display distribution is the small “keyhole” the CDU provides the pilot to peek in on the workings of the FMS. It has already been mentioned that the CDUs have in the neighborhood of 30 pages, only one which may be displayed at a time. In total, the A320 has about 150 editable fields and menus options (excluding those that merely aid navigation in the CDU itself), of which generally less than 10 are visible at any one time.

Each of these editable fields or menus options is approximately equal to a single physical control (a knob or selector switch respectively) The A320 has about 250 nonredundant physical controls within reach of the pilots (excluding the CDU). Thus, over one third of the “controls” for the aircraft (about 150 out of 400) are concentrated in one small 10" × 6" panel, while all remaining controls get relatively luxurious quantities of panel space.

This promotes system insulation and its contribution to lack of mode awareness and diagnosis difficulty, as most of the FMS settings are necessarily out of view at any given moment. Because settings that affect the same parameter of performance may be on different pages of the CDU, the small keyhole also effectively increases the distribution of controls and displays. This further contributes to a lack of mode awareness, programming errors, and diagnosis difficulty.

In addition to the above contribution to AFS problems, the small keyhole afforded by the CDU adds to programming difficulty because it forces the pilot to electronically navigate the pages of the CDU to get to displays and controls.
Unlike physical controls, it is not enough for the pilot to know where a display or control is. She or he must also know how to get there. Maintaining this knowledge imposes is an additional burden on training and an additional factor on workload.

Even if the knowledge is easily retrieved, at the very least, the keyhole problem slows a pilot’s ability to make an input compared to physical controls because some of the keypresses are merely to get to the right place to make the input. For example, to execute a “direct to” (take a new track toward a specific waypoint), a pilot will typically execute nine discrete actions on the Boeing 747-400 FMS, up to four of which are purely for navigating within the CDU. Some have noted that AFS-equipped aircraft are slower to respond to ATC request because of the many steps needed to reprogram an FMS (Bluecoat, 1997b).

It has been said that the FMS is not just a component of the aircraft – it is the aircraft (Abbott, 1997). If so, then the pilot is forced to view the aircraft through a small $24 \times 14$ character display (along with some things on the ND and PFD).

**PATHS TOWARD SOLUTIONS**

*Principles*

There is a general need for development of improved displays and controls for AFSs (FAA, 1996). Such a revised user interface for the AFS can address the above sources of automation problems by having the following characteristics:

1. Have a consistent user interface, internally and across aircraft.
2. Have fewer modes, or at least have more organized modes.
3. Provide detailed information about the aircraft state when needed.
4. Be physically and consistently organized throughout the cockpit with respective to the physical controls and displays.
5. Maximize the simultaneous availability of important displays and controls, and minimize the need to navigate the system itself.

Above and beyond that, such a new user interface for the AFS should also be designed consistent with the following characteristics, each largely derived from standard human factors principles:

1. If automation is to be used at all during times of high workload, then the user interface must be optimized for fast, easy input of changes especially from ATC clearances. This will reduce programming errors and distraction by time and effort spent on diagnosis. Specifically:
(a) Flightplans should be editable with a minimum of keystrokes. Pilots should be able to change any detail of the flight plan without having to re-enter whole portions of it. The capacity to quickly change approaches is especially important. An Undo capacity should be provided.

(b) Language and syntax must be intuitive to pilots and consistent with format given by ATC (FAA, 1995; Proctor, 1997; Riley et al., 1998).

(2) Automation needs a clearer display of its state, current and future. All parameters and values of the state must be displayed. Changes in state (e.g. mode) ideally should be made clear to pilots when they occur. Automation also needs to clearly alert pilots early if a program state cannot be reached and what automation intends to do about it (e.g. what mode it will fall into). These two alerts must be clearly distinct from each other. Automation needs this to prevent programming errors and loss of mode awareness, and to facilitate diagnosis of problems. More specifically, the AFS should:

(a) Use plain language in displays rather than symbols (e.g. “Before” not “-”).
(b) Provide clear display of default values in the same place where other values are entered.
(c) Avoid, as much as possible, multi-function controls (Sherry et al., 2001); where they cannot be avoided, at least dynamically change the label on a multi-function control to represent its current function.
(d) Show the effects of input at the point the input is made. Annunciate mode changes where the pilot is looking when making the mode change. Warn of flight plan automatic overrides at the time a change is programmed.

(3) Either the effects of user action on the modes has to be consistent with the pilots’ model, or the pilots have to be trained on the model the automation follows. Displays and controls should be designed to intuitively communicate the model. The model must be internally consistent as well. This will allow users to anticipate a mode change before it appears on the display, and to more quickly guess the reason automation performed the way it did.

(4) Any air transport user interface must be compatible with two-person crew, facilitating verification procedures and intercommunication of actions and intentions.

The new interface will also have to be designed to fit well in the future of information-age flying as well as present-day flying. Specifically, it will have to be support required navigation performance (RNP) navigation, ATC datalink
transmission of clearances, and Free Flight. It should also be integrated with digital approach plates, en route charts, and other electronic references.

Technologies

There are multiple technological paths that may help achieve such a new user interface, each needing to be explored before being considered for implementation. Some paths may prove to be infeasible or unhelpful, but by sensibly combining the successful paths, a new cockpit will emerge tailored to modern automated flight. This section gives some candidate technological paths.

Interactive EFIS

Modern glass cockpits already provide pilots with a graphical representation of navigation information in the form of the ND of the EFIS. With an interactive EFIS, pilots would be able to control navigation, as well as view it, through direct manipulation of the symbols of the ND (Holahan, 1996). For example, pilots could select a waypoint and then, through a pop-up menu, execute a Direct To towards it. Essentially, the pilots would have a graphical user interface (GUI) to the FMS (see Fig. 3).

An interactive EFIS would have the following advantages:

1. Navigation inputs remain in the visual channel consistent with the pilot’s mental visualization, reducing the chance of error in translating the visualization to a verbal/alphanumeric channel (Wickens, 1984).
2. The geographical path of the programmed course is apparent before the pilot even finishes entering it, reducing the reliance of cross-checking with the ND as must be done now.4
3. Certain actions can be completed faster using an object-action direct manipulation syntax made possible with an interactive EFIS. For example, a Direct To could be executed in one to four steps, rather than the nine it typically takes now with the text-based CDU.

A GUI would require different input devices than currently exist for AFSs. Some sort of pointer device is needed, one that remains controllable when turbulence or other disturbances are encountered. Pointer devices are already in use on some aircraft. The Boeing 777 has a touchpad for controlling some systems and various military aircraft have pointer devices. In addition to the touchpad, a trackball, or pedestal-mounted sidestick are possibilities. The pointer device would have to have integrated buttons to select, execute, and cancel, perhaps with a separate button for each. Possibly, the pointer would also have some device, such as a thumbwheel, for entering numeric field values (e.g.
headings, altitudes, speeds, frequencies). Alternatively, a logical “spinner” or “slider” control could be used on the display itself, as is the practice on microcomputer GUIs.

For alphanumeric field input, voice recognition could be used. Voice recognition of conversational speech is still an emerging technology, but state of the art voice recognition may be able support the limited vocabulary necessary to operate an FMS, being the International Civil Aviation Organization (ICAO) phonetics for the digits 0–9 and the letters A–Z, and perhaps some basic commands. Pilots already frequently communicate through ICAO phonetics and use headsets which can mount boom microphones, making them ideal candidates for today’s voice recognition technology. Voice recognition eliminates the need for a keypad, thus saving panel space.

Fig. 3. A Hypothetical Interactive EFIS Display with one Waypoint Selected.
(although a slide-out backup keypad may be desirable) and allows for more head-up time. Pilots may even be able to phonetically spell something faster than they can enter it on a keypad. Finally, using a voice recognition interface provides an opportunity for a crew to cross-check each other (reducing system insulation). One pilot will hear the input made by the other pilot, something which does not naturally occur with a keypad interface.

Voice recognition would not be used instead of direct manipulation of the interactive EFIS. Rather, voice recognition and direct manipulation would be used together, each for the cases for which it is best suited. For example, to enter a waypoint within an existing leg, pilots may select the leg from the ND, select “Insert Waypt” from a pop-up menu, and then verbally spell the name of the desired waypoint, keying the microphone to the FMS with the “select” button on the pointer device.

Ideally, the interactive EFIS and voice recognition would be sufficiently advanced to subsume all of the aviating and navigating control functions of the CDU, or even eliminate the CDU entirely. With an interactive EFIS, non-FMS functions may also be controlled from it, in particular the functions of the MCP are a desirable candidate, either eliminating the MCP or rendering it a backup. This would result in all aircraft AFS control and display being co-located, reducing the distribution of displays and controls. Change in the flight path would be in foveal view when pilots change the operating mode.

Ideally, the EFIS would provide the following displays, either as selectable layers within a single display such as the ND, or as alternative displays within the same panel area:

1. Navigation information, traffic, weather, terrain, and electronic charts.
2. Programmed, alternative, and predicted aircraft path.
3. Mode transitions along the path showing the operating modes for each leg (including present position), perhaps displayed in the margins (Vakil et al., 1995).
4. A hierarchical list of the flight plan showing waypoints within airways, approaches, flight phases and so on, with a capacity to interactively expand a waypoint or leg for more details.
5. A graphical display of the flight plan altitude profile (a “vertical navigation display”), with direct manipulation of the floor and ceiling (Oman, Kendra, Hayashi, Stearns & Bürki-Cohen, 2000; Rosay, 2000; Vakil et al., 1996).
6. Perhaps a similar display showing the aircraft’s speed and speed restrictions among waypoints.

Simply putting a GUI on the flightdeck will not necessarily guarantee an improvement. A GUI that simply replicates the CDU with proportional font and dialogue
boxes, instead of pages will not result in a substantial improvement (Abbott, 1995). Instead, an interactive EFIS would have to be carefully developed, taking into account the entire AFS and using user interface techniques such as voice recognition and direct manipulation where they will be the most effective.

Multi-channel Mode Feedback

The usefulness of sophisticated operating mode displays should be investigated to improve mode awareness. Already, air transports display on the PFD the current and next anticipated modes for vertical, lateral, and thrust dimensions. It has already been described how, through the margins of the ND, the AFS can display not only what the next mode is, but when it is expected to engage. Related to this would be some sort of display indicating when an envelope is about to be breached, particularly if it means the automation will switch off. This can be some display of distance to the edge of an envelope or merely an annunciation when the aircraft is within certain bounds of the envelope.

A more sophisticated display would also give more details about a particular mode, namely some subset of the parameter, reactor, source, and target. The presence of any sub-modes may also be important information. Tabular display of some of this information is useful to pilots for the vertical navigation modes (Vakil et al., 1995).

For reducing system insulation, perhaps it is more important to provide immediate feedback on any changes in aircraft behavior, rather than to identify the current and next mode. Current AFSs indicate mode changes by the change in an annunciator on the PFD. While the PFD is part of the regular scan, indicating changes through the visual channel will inherently have lag. Empirical study may test if the auditory channel can be used more (Monk, 1986). At a minimum, the AFS can orient the pilot to check the PFD by giving an unobtrusive sound (such as the “click” sound used in A300s). At the other extreme, the AFS can call out the new mode by name with a synthesized voice, although to minimize distraction this perhaps should be reserved for abnormal circumstances (e.g. envelope breech).

As an additional or alternative role for the auditory channel, it may be argued that immediate feedback of changes in the operating mode are actually not as important as immediate feedback on changes in aircraft behavior at a more fundamental level. This implies that a greater benefit may be realized by providing redundant auditory or tactile displays of selected PFD indicators. For example, a pilot should know immediately if the aircraft stops flying level and begins to descend. If necessary, the pilot can then check other visual displays to determine the reason for the change to descent if it is not known. Given that modern cockpits already have many sound displays, and the need to be able
to monitor ATC communications effectively, efforts should be made to minimize additional auditory displays. Tactile displays are generally limited in the dimensions they can communicate. A display of the basic elements of aircraft behavior (e.g. changes in vertical speed, roll, and airspeed) reduces the number of unique features to encode, certainly less than the 20 to 80 needed to encode each operating mode.

The “Real Estate” Issue

As may be gathered by the suggestion for an interactive EFIS and expanded operating mode displays, finding space for these components on the flight deck may be problematic. Ironically, greater automation does not mean less displays and controls, as there must now be displays and controls for the automation in addition to backup displays and controls for the manual systems. The most obvious solution is to simplify the AFS, removing operating modes and consolidating features into more general internally consistent functions. However, current trends are to increase the features on AFS and related systems, including:

2. ATC datalink.
3. Cockpit display of traffic information (CDTI).
4. Electronic charts and checklists.
5. Time of arrival control (TOAC).
6. Synthetic and enhanced vision systems.

All the above features will require additional displays and controls somewhere in the cockpit. It appears inevitable that pilots will have more to monitor and control in future aircraft.

To prevent further aggravation of the keyhole problem through devices such as the CDU, study is needed to find better ways to have more controls and displays to exist in the same size cockpit. Among the technology available for investigation is the following:

• **Head-up Displays (HUD).** The greatest benefit of a HUD PFD may not be for directing the pilot’s eyes (if not his or her attention) out the window, but rather in allowing more panel space to be available for something like an interactive EFIS, vertical navigation displays, and so on. Given that HUDs symbology is only visible if the pilot’s eyes are with a prescribed space (Newman, 1995), this effort to increase space is most practical for short durations of flight, such as during approach and landing.
• **Vertical Information and the ND.** The high rates of incidents and accidents associated with vertical navigation in automated aircraft suggest that pilots need more vertical flight path information in the cockpit (Palmer et al., 1993; Vakil et al., 1995). Given the competition for panel space, the relative merits of a dedicated flight path profile display needs to be compared with placing richer vertical path information on the plan-view ND (which would, however, add to its clutter). Possibly each is better under different circumstances and a capacity to switch the display is needed.

• **Multi-function Displays.** It is apparent that pilots will need to call up different displays for different circumstances even more in the future. Pilots will need a standardized means of accessing different displays and flexibly but quickly selecting a subset of the total number of displays to show at once. Doing so through physical switches, as is done today, may not be adequate on future flight decks. This may be something best done through logical controls. It is anticipated that a “windows” approach may be too awkward, so alternatives must be proposed and explored.

• **Automatic Display Switching.** With multiple displays, the pilot must now spend time and effort managing the displays, adding to workload. This is seen today in CDUs where pilots must hunt through multiple pages to find a target field. Perhaps the automation can aid the pilot in this task by intelligently changing the displays in anticipation of the pilot’s needs (e.g. switching to shorter range plan views as the aircraft approaches a destination, or changing the plan view to a profile view after the final approach fix). The benefits of this sort of automation have to be weighed against the potential problems associated with any automated or programmed changes as described earlier.

• **Hierarchical User Interface.** A systematic hierarchical user interface may best allow displays to share the same panel space. The most general and most critical information is displayed at the top level. Pilots then have the ability to “drill down” quickly to lower more detailed levels for the purpose of atypical programming or diagnosis.

*Full Cockpit Use*

While the FMS integrates all major aircraft systems, pilot access to it is largely concentrated on the CDU. It may be more effective to make the entire cockpit the user interface to the FMS, with variable displays and logical controls for the FMS located by the physical controls for lower levels of automation. For example:
(1) All navigation functions are centralized on the ND.
(2) The equivalent of the radio tuning page collections in the CDU would be located on the radio console.
(3) The progress page would be on the MCP.

To the degree the FMS controls and displays, whether logical or physical, combine with the non-FMS controls and displays, effectively greater panel space becomes available for the FMS fields while reducing the physical distribution of controls and displays for a given functional area. Rather than looking though the CDU keyhole to see the FMS, all panel spaces are available.

*On-Line Help*

Unless there is a radical change in the training procedures, it has to be assumed that some on-the-line learning of the less critical aspects of AFSs will continue. Thus, it would seem wise to provide some sort of context sensitive on-line help in the AFS to supplement any manuals. At the least, pilots should able to quickly ask the AFS “why is it doing that” (facilitating diagnosis) and maybe even “how do I get it to do what I want” (facilitating less error-prone programming). Such on-line help is not that far removed from checklists and manuals pilots regularly carry aboard aircraft.

*The Next Generation Cockpit*

There is a need to develop the “Macintosh” of flight automation, not in the sense that it would necessarily look and feel like the Apple Macintosh itself, but in the sense that it would provide an advanced, standardized, integrated, intuitive, and powerful grammar of human-machine interaction for AFSs. In carrying out this development, we may some day look back to the first generation glass cockpits, with their inconsistent, difficult, and disjoint interfaces, as a transitional stage to a more mature form of pilot control of an automated aircraft.

*NOTES*

1. Usually referred to as “Auto Thrust” in the Airbus A320.
2. Sometimes a CDU are incorrectly said to have a command line interface to distinguish it from a graphical user interface. Technically, a command line interface is one in which the system presents minimal prompting and the user would type in commands
one letter at a time using an alphanumeric keypad. If the CDU were command line, the
user would have to type something like “ALT BAKER 13000+” to set an altitude
constraint for the Baker waypoint. In a form interface, the system presents physically
distributed and labeled data fields for input. The user enters an altitude constraint into
a blank provided for it. In a menu interface, commands are selected from a list of choices
presented on the screen.

3. There may be up to 24 editable fields per page as CDU’s often allow users to select
and enter two fields at once with one LSK, separating the two field values with a slash.

4. It can be imagined that the waypoint entry errors made by the pilots of American
Flight 965 to Cali on 12/20/95 would have been vastly less likely if the pilots had to
select the waypoints from the ND itself.

REFERENCES

Proceeding of the Eighth International Symposium on Aviation Psychology, Columbus, OH,
April.
Avsoft Systems, Aurora, CO.
Bluecoat (1997b). FMS aircraft are slow to execute. Thread beginning November 14.
Buffer, B. (1999). Big boeing FMC user’s guide (2nd revision). Kingwood, TX: Leading Edge
Libraries.
Barbara: Aegean.
awareness difficulty for pilots of glass cockpit aircraft. IEEE Transactions on Systems, Man,
OSUasaf.html
Federal Aviation Administration. (1996). Federal Aviation Administration human factors team
report on: The interfaces between flightcrews and modern flight deck systems. Washington,
D.C.: Author.
Holahan, J. (1996). LCDs, mice on the flight deck! Digital engine to drive Primus avionics into
the dawn of the next century. Aviation International News, November 1.
Hughes, D. (1997). FAA inspection system works, but challenges loom. Aviation Week and Space
Technology, August 18, 60–62.
May 9, 31–32.
Monk, A. (1986). Mode errors: A user-centered analysis and some preventative measures using
gation displays: Pilot performance and workload during simulated constant-angle-of-descent

ABBREVIATIONS AND ACRONYMS

A/P Autopilot
A/T Auto Throttle or Auto Thrust
ACARS Aircraft Communications Addressing and Reporting System
AFS Automated Flight System
ALT (1) Altitude, (2) Altitude Hold Mode
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT CAP</td>
<td>Altitude Capture Mode (McDonnell Douglas)</td>
</tr>
<tr>
<td>ALT*</td>
<td>Altitude Capture Mode (Airbus)</td>
</tr>
<tr>
<td>ALT/HOLD</td>
<td>Altitude Hold Mode (McDonnell Douglas)</td>
</tr>
<tr>
<td>APP</td>
<td>Approach Mode</td>
</tr>
<tr>
<td>ATC</td>
<td>Air Traffic Control</td>
</tr>
<tr>
<td>CAS</td>
<td>Calculated Air Speed</td>
</tr>
<tr>
<td>CDU</td>
<td>Control Display Unit</td>
</tr>
<tr>
<td>CDTI</td>
<td>Cockpit Display of Traffic Information</td>
</tr>
<tr>
<td>CI</td>
<td>Cost Index</td>
</tr>
<tr>
<td>CLAMP</td>
<td>Throttle Clamp Mode (McDonnell Douglas)</td>
</tr>
<tr>
<td>CLB</td>
<td>Climb Mode (Airbus)</td>
</tr>
<tr>
<td>DES</td>
<td>Descend Mode (Airbus)</td>
</tr>
<tr>
<td>DME</td>
<td>Distance Measuring Equipment</td>
</tr>
<tr>
<td>ECON</td>
<td>Economy Mode</td>
</tr>
<tr>
<td>EFIS</td>
<td>Electronic Flight Instrument System</td>
</tr>
<tr>
<td>EPR/LIM</td>
<td>Engine Pressure Ratio Limit Mode (McDonnell Douglas)</td>
</tr>
<tr>
<td>EXP CLB</td>
<td>Expedite Climb Mode (Airbus)</td>
</tr>
<tr>
<td>EXP DES</td>
<td>Expedite Descent Mode (Airbus)</td>
</tr>
<tr>
<td>FAA</td>
<td>Federal Aviation Administration</td>
</tr>
<tr>
<td>FCU</td>
<td>Flight Control Unit (Airbus)</td>
</tr>
<tr>
<td>FD</td>
<td>Flight Director</td>
</tr>
<tr>
<td>FLCH SPD</td>
<td>Flight Level Change – Speed Mode (McDonnell Douglas)</td>
</tr>
<tr>
<td>FLX</td>
<td>Flex Take-off Mode (Airbus)</td>
</tr>
<tr>
<td>FMC</td>
<td>Flight Management Computer</td>
</tr>
<tr>
<td>FMCS</td>
<td>Flight Management Computer System (Smith Industries, B737)</td>
</tr>
<tr>
<td>FMGC</td>
<td>Flight Management Guidance Computer (Airbus)</td>
</tr>
<tr>
<td>FMS</td>
<td>Flight Management System</td>
</tr>
<tr>
<td>FPA</td>
<td>Flight Path Angle</td>
</tr>
<tr>
<td>Fpm</td>
<td>Feet Per Minute</td>
</tr>
<tr>
<td>F-PLN</td>
<td>Flight Plan</td>
</tr>
<tr>
<td>FY</td>
<td>Fiscal Year</td>
</tr>
<tr>
<td>G/S</td>
<td>(1) Glideslope, (2) Glideslope Mode</td>
</tr>
<tr>
<td>GA</td>
<td>(1) General Aviation, (2) Go Around (Boeing)</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HDG</td>
<td>Heading Select Mode</td>
</tr>
<tr>
<td>HDG HOLD</td>
<td>Heading Mode – Holding Pattern</td>
</tr>
<tr>
<td>HDG SEL</td>
<td>Heading Select Mode (Universal)</td>
</tr>
<tr>
<td>HOLD</td>
<td>Holding Pattern Mode</td>
</tr>
</tbody>
</table>
HUD  Head-up Displays
ICAO  International Civil Aviation Organization
ILS  Instrument Landing System
LAND  Auto Land Mode
LNAV  Lateral Navigation
LOC  Localizer Mode
LOC*  Localizer Capture Mode (Airbus)
LSK  Line Select Keys
MACH  Mach Speed Mode
MCDU  Multifunction Control Display Unit
MCP  Mode Control Panel (Boeing)
MCT  Maximum Continuous Thrust
MSL  Mean Sea Level (i.e. the altitude above it)
NAV  (1) Navigation, (2) Lateral Navigation Mode
ND  Navigation Display
OP CLB  Open Climb
OP DES  Open Descent
PERF  Performance Page (of an FMS)
PFD  Primary Flight Display
RNP  Required Navigation Precision
RWY  Runway Mode
SID  Standard Instrument Departure
SP  Scratch Pad (on a CDU)
SPD  Speed Mode
SRS  Speed Reference Scale Mode (Airbus)
STARS  Standard Terminal Arrival Routes
THR  Thrust Mode
THR REF  Thrust Reference Mode (McDonnell Douglas)
TOAC  Time Of Arrival Control
TOGA  Take-off – Go Around Mode
TOGA LK  Take-off – Go Around Lock (Airbus)
TRK  Track
V/S  (1) Vertical Speed, (2) Vertical Speed Mode
VERT/SPD  Vertical Navigation – Speed Mode
VHF  Very High Frequency
VNAV  Vertical Navigation Mode
VNAV ALT  Vertical Navigation – Altitude Mode
VNAV PTH  Vertical Navigation – Path Mode
VNAV SPD  Vertical Navigation – Speed Mode
VOR  VHF Omni-directional Ranging
This Page Intentionally Left Blank
8. AUTOMATION ISSUES IN REGIONAL AIRLINE OPERATIONS

Beth Lyall, Greg Harron and Jennifer Wilson

INTRODUCTION

The purpose of this chapter is to describe the work we have been conducting to understand regional airlines operations and the availability and use of automation. Previous automation studies have addressed automation use in large transport aircraft by pilots in the major airlines. In this work, we are focusing on describing automation use by the regional airlines and how it is different from what has been found studying the large transport airlines. First will be presented general information that we have developed about pilot tasks and automation. Next, information and data will be described and used to make general comparisons between the regional operations and transport operations and their use of automation. Finally, a summary of the lessons we have learned so far in this work will be presented.

PILOT TASKS

In considering the effects of any technology or other change in the flight deck it is important to address the change in light of the possible effects of the change on the structure and performance of the pilot tasks. The very general pilot tasks of aviate, navigate, communicate, and manage systems have been commonly used to describe the scope and priorities of pilot responsibilities. We have
modified the fourth task to be stated “manage flight and systems” to more clearly communicate the addition of the strategic nature of that task and its distinction from the tactical nature of the other three tasks. We have found that breaking down the four general tasks to one more level helps as we use them to understand the implications of the addition of automation in the flight deck. This breakdown is presented in Table 1.

We believe that the task list in Table 1 is fairly complete at this general level. Within the four general tasks of aviate, navigate, communicate, and manage flight and systems it can be seen that the task list includes four types of subordinate tasks: control, configure, monitor, and manage. We have taken the approach that the four general tasks present the objective to be achieved by their subordinate tasks. The “aviate” tasks include all requirements of the pilot to maintain flight. The “navigate” tasks include all those required for directing the aircraft from its origin to its destination. The “communicate” tasks

### Table 1. Pilot Tasks.

<table>
<thead>
<tr>
<th>General Task</th>
<th>Subordinate Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aviate</td>
<td>Control airplane to maintain flight (airspeed, altitude, attitude)</td>
</tr>
<tr>
<td></td>
<td>Monitor flight parameters (airspeed, altitude, attitude)</td>
</tr>
<tr>
<td></td>
<td>Configure flight surfaces</td>
</tr>
<tr>
<td></td>
<td>Configure flight guidance system (autopilot) for control of airplane</td>
</tr>
<tr>
<td></td>
<td>Monitor flight guidance system (autopilot) control of airplane</td>
</tr>
<tr>
<td>Navigate</td>
<td>Control airplane direction in flight in accordance with flight plan</td>
</tr>
<tr>
<td></td>
<td>Control airplane direction on the ground</td>
</tr>
<tr>
<td></td>
<td>Configure flight guidance system and/or FMS for navigation of airplane</td>
</tr>
<tr>
<td></td>
<td>Monitor flight guidance system and/or FMS navigation of airplane</td>
</tr>
<tr>
<td>Communicate</td>
<td>Configure communication system</td>
</tr>
<tr>
<td></td>
<td>Control (initiate and respond to) communications</td>
</tr>
<tr>
<td></td>
<td>Monitor communications</td>
</tr>
<tr>
<td>Manage flight and systems</td>
<td>Manage flight (prioritize resources and systems; develop and modify flight plan)</td>
</tr>
<tr>
<td></td>
<td>Configure systems</td>
</tr>
<tr>
<td></td>
<td>Manage systems</td>
</tr>
<tr>
<td></td>
<td>Monitor systems</td>
</tr>
<tr>
<td></td>
<td>Monitor flight path, fuel consumption, and ATC clearance compliance</td>
</tr>
<tr>
<td></td>
<td>Monitor traffic, weather, terrain, and obstacles</td>
</tr>
</tbody>
</table>
include requests for, and reception of, instructions and information with air
traffic services, the airline (dispatch, operations, maintenance, and ground
crews), other pilots, and cabin crewmembers. The “manage flight and systems”
tasks in general cover the planning and monitoring of the flight, and the manage-
ment and monitoring of systems and resources.

Each of the subordinate task statements begins with the task type and then
describes the focus of the task within its general objective. Notice that control,
configure, and monitor tasks are associated with all four of the general
tasks. Only the subordinate “manage” tasks are exclusive to their general task
objective because it captures the strategic components of the general task. The
subordinate tasks seem to be more useful in directing our work related to
automation use because they are described at a level that is more directly related
to the descriptions of automation functions. The next section will build on this
information when defining the categories of automation.

CATEGORIES OF AUTOMATION

We have found that the definition of automation is most useful when based on
function, rather than on design implementation. We use a very basic definition
of automation: a system function is automated if it accomplishes something that
the pilot would otherwise have to accomplish. Many general terms have
been used to refer to flight deck automation when it is being studied: glass,
modern, automated. However, none of these terms is very useful because it is
not descriptive of the functions that have been automated. Rather than
addressing automation as a broad concept, it is effective to consider categories
of flight deck automation that are based on the objectives or functions of the
automation components. Three categories of automation have been defined:
control automation, information automation, and management automation
(Fadden, 1990; Billings, 1997). These three categories of automation are very
different from each other in the automated functions they provide, and
separating them facilitates focusing on the specific functions provided to the
pilots and how their tasks may change. The three categories are described in
detail in the following subsections.

Control Automation

The purpose of control automation is to accomplish a control or configuring
task that the pilot would otherwise perform. The autopilot is the most often
studied and discussed instance of control automation. Other examples are
auto-throttle (auto-feather on propeller driven aircraft), auto-land, auto-trim,
auto-brake, and automated functions that set system parameters. The dichotomy most often used for this category of automation is automated vs. manual. The pilot would manually accomplish the task if it was not being accomplished by an automated function.

Because the training and practice of these manual tasks lead the pilots to gain skills associated with performing them, the loss of the manual skill has been a concern raised in many discussions of automation use. In particular, the loss of manual-, or hand-, flying skills have been much-visited topics. Theoretically, it would be possible to have some sort of skill loss with continued use of any control automation functions because their use would reduce the amount of practice that the pilot gets on the psycho-motor skill affected by the automated function. However, loss of skill may also occur for pilot performance associated with functions in the other two categories of automation as will be described in the following subsections.

**Information Automation**

The purpose of information automation is to process information that the pilot would otherwise have to process or transform, or to integrate several sources of information or data into one picture that the pilot would otherwise have to integrate. The most common example of information automation is the map display that integrates many sources of information and presents the picture of the aircraft location in relation to the programmed flight path and the surrounding geography (and possibly other selected identifiers such as navigation aids or airports). The moving map display has been so heavily associated with the discussion of automation that many use the term “glass cockpit” to refer to a flight deck that is highly automated.

Information automation was made possible when digital information became available for use in the design of flight deck displays. With the digital revolution came the capability to process raw information and transform it to a form more useful to the pilots. A simple example of this type of processed information is the digital display of fuel quantity instead of the raw information presented on an analog display of relative fuel quantity. This is information automation because the system processes the raw information and transforms it in a way that the pilot would otherwise have to do to accomplish certain tasks such as deciding how much additional fuel to add. This simple example illustrates the basic concept of automation: that to be considered automation, it must perform a function that the pilot would otherwise perform.

The flight director is the information automation component that is most available on flight decks. It presents integrated guidance information that
can be used by the pilot or the autopilot during control of the aircraft. Other examples of information automation are the primary flight displays, electronic engine instruments, ground proximity warning systems, traffic collision and avoidance systems, and electronic engine instruments. The processed or integrated information presented by information automation may contribute to the pilot performance of any of the monitoring, control, or management tasks because they all require information to be accomplished.

Similar to control automation, the use of information automation may result in pilot skill loss. The skills associated with tasks that require the processing and integrating of information may be degraded if the pilots do not practice them at a certain level of frequency. We hope to pursue this question in the future to determine how vulnerable these skills may be to degradation.

**Management Automation**

Where control automation focuses solely on, and information automation primarily focuses on, helping the pilot with tactical tasks, management automation can also focus on the performance of strategic tasks. Management automation performs functions related to strategic planning and management of the aircraft operation that the pilot would otherwise perform. Management automation functions provide resources for performance of the strategic tasks associated with flight and system management, such as deciding on future waypoints to include along the flight path, determining the best altitude to use to optimize fuel consumption, or choosing the most appropriate alternate airport. The pilot skills associated with the tasks related to the functions of management automation are cognitive skills such as planning, decision-making, or problem solving. As with the skills related to the other two categories of automation, it is possible for these skills to degrade if they are not exercised frequently enough to maintain them. Therefore, it is possible for the over-use of management automation to result in skill degradation just as with the other automation categories. This concept has not yet been explored specifically by research to understand how to reduce the skill degradation effects while maintaining the benefits to accomplishing such complex tasks.

**AUTOMATION AND REGIONAL AIRLINES**

One point of interest in understanding the regional airline population is how they differ from the major transport airlines. This is of interest because most research to date has focused on the transport operations and it would be useful
to know the degree to which that research can be applied to the regional airlines. We have been using the automation categories to aid our understanding of these differences.

Our findings indicate that the regional airlines differ from the transport airlines in many ways. The regional airlines are not just smaller transport airlines, but they have many unique characteristics that will be important to consider when addressing their use of automation. One important finding is that the pilot populations of, and aircraft flown by, the regional airlines vary much more than these same characteristics of the large transport airlines. To describe these characteristics and make a comparison between the regional and transport airlines, members of the Regional Airline Association (RAA) were used to represent the regional airlines, and members of the Air Transport Association (ATA) were used to represent the major airlines. Table 2 presents a general comparison. It can be seen that there are many more regional airlines than ATA airlines and that the regional airlines fly many more types of aircraft. However, the number of aircraft and departures is lower for the regional airlines. We supplemented the demographic information from the RAA and ATA with selected interview information. Another difference identified is related to the experience and turnover of their pilots: RAA pilots are typically less experienced than ATA pilots, and the pilot turnover rate within regional airlines is usually much higher than for the major carriers.

Figure 1 presents information about the number of aircraft types and our initial findings about the automation that is typically available for the aircraft flown by the ATA and RAA airlines. Figure 2 presents these same data as percentages. The automation available for the different aircraft types was identified through many sources such as interviews, aircraft flight deck photographs, and descriptions of aircraft specifications. It is particularly difficult to obtain reliable information for all of the RAA aircraft types. We are still identifying good sources for what is a typical configuration for some of the aircraft. In the future we hope to be able to

<table>
<thead>
<tr>
<th></th>
<th>RAA Members</th>
<th>ATA Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of airlines</td>
<td>97</td>
<td>28</td>
</tr>
<tr>
<td>Number of aircraft types</td>
<td>98</td>
<td>26</td>
</tr>
<tr>
<td>Number of aircraft</td>
<td>4,308</td>
<td>5,647</td>
</tr>
<tr>
<td>Number 2000 departures</td>
<td>4.46 million</td>
<td>8.99 million</td>
</tr>
</tbody>
</table>

*Note: Information obtained from ATA and RAA websites, October 2001.*
Fig. 1. Number of Aircraft in ATA and RAA Fleets with Various Types of Automation Available.

Note: The analysis included 26 aircraft types from the ATA fleet and 48 aircraft types from the RAA fleet. Aircraft types that represent fewer than 20 aircraft within their respective fleet (ATA or RAA) were not included.
Fig. 2. Percentage of Aircraft Types in ATA and RAA Fleets with Various Types of Automation Available.

Note: The analysis included 26 aircraft types from the ATA fleet and 48 aircraft types from the RAA fleet. Aircraft types that represent fewer than 20 actual aircraft within their respective fleet (ATA and RAA) were not included.
gather information from the regional airlines about the particular automation systems available on the aircraft they fly. These descriptions will be more valuable than those of typical configurations for understanding training needs for automation and other questions. With the lower pilot experience and higher pilot turnover rate of the regional airlines, the automation training issues could be even more daunting than for the major carriers.

Notice in the figures that we included both auto-feather and auto-throttle under control automation. We did this because of the large numbers of propeller-driven aircraft flown by the regional airlines. This is evident in the data through the many more aircraft with auto-feather in the regional airlines and the many more with auto-throttles in the transport airlines. Because of the large number of aircraft types flown by RAA members, all other indications in Fig. 1, except FMS, show many more RAA aircraft types represented. However, Fig. 2 shows that the percentages are similar for everything except auto-feather and auto-throttle (for reasons already presented) and for FMS. About 63% of the ATA member aircraft are equipped with an FMS where about 38% of the RAA aircraft have an FMS. In interviews with some of the regional airline training personnel we learned that many of the airlines are trying to increase their number of new regional jet aircraft, which are equipped with an FMS, and reducing their number of older propeller-driven aircraft. However, even though the regional airlines are transitioning to aircraft more like those flown by the transport airlines, they are facing many additional challenges related to implementing the automated aircraft because of the differences in their pilot background, experience, and turnover.

**ISSUES RELATED TO USING AUTOMATION**

To identify recurring pilot-automation interaction problems, a set of flight deck automation issues has previously been developed based primarily on research and information from major transport airline operations (Funk & Lyall, 1997). As part of this research, a review of incident reports for major transport airlines was conducted to identify the prevalence of these flight deck automation issues. For comparison to the major transport airlines, this set of automation issues has also been used in a review of regional airline incident reports to identify the flight deck automation issues that exist for that population. For each of these incident analyses, incidents from the Aviation Safety Reporting System (ASRS) Incident Database were reviewed. In the regional airline incident analysis, incident reports submitted to ASRS between January 1994 and September 1998 were reviewed. Automation issues were identified in 84 regional airline incident reports. In the major transport airline analysis, incident reports
submitted between 1988 and 1996 were reviewed. Automation issues were identified in 284 major transport airline incident reports. For more information about the development of these automation issues and other details related to this research refer to the Flight Deck Automation Issues (FDAI) website (www.FlightDeckAutomation.com). Even though these analyses are not equivalent, the data provide a good starting point to understand the differences or similarities in automation issues. In our further work we will update the transport incident analysis and expand the regional incident analysis to make the comparison more effective.

Table 3 compares the regional airline incident analysis results with those of the major transport incident analysis. The table presents a list of the most prevalent flight deck automation issues identified in the incident reports. The issues are presented in the order of frequency of occurrence in the regional airline incident review. The rank order, percentage of incidents, and number of incidents for each issue is listed for the regional airlines and major transport airlines, respectively. It can be seen that, in general, the more frequently cited issues from regional airline incidents are also the more frequently cited in the major transport incidents.

The two most commonly occurring issues were identical for the regional airlines and the transport airlines. The top ranking issue for both analyses was “Pilots may be overconfident in automation,” and the second ranking issue was “Automation behavior may be unexpected and unexplained.” The expanded description of the issue about pilot overconfidence in automation describes occurrences in which pilots may become complacent and fail to exercise appropriate vigilance because they are overconfident in, and uncritical of, automation, sometimes to the extent that the pilot abdicates responsibility to automation. The expanded description of the second issue for both populations says that it addresses instances when automation may have performed in a manner that was unintended, unexpected, and perhaps unexplainable by pilots, possibly creating confusion, increasing pilot workload to compensate, and at times, leading to unsafe conditions.

The overall picture created by these incident analyses reveals that the flight deck automation issues being written about by pilots submitting incident reports are very similar in the two populations. However, upon closer inspection of the data, some differences between the populations may be identified.

The issue of inadequate understanding of automation seems to be more prevalent in the regional airline population. Inadequate understanding of automation was found to be a relevant issue in just over 14% of the regional airline incident reports compared to only about one percent of the major transport incident reports (see Table 3). This discrepancy in the two populations may be partially explained by the demographics of the regional airline population
discussed earlier in the chapter. The regional airline pilots are typically less experienced, have a higher turnover rate, and may fly a wider range of aircraft types as compared to the major transport airlines. Each of these factors may affect the regional airline pilot’s understanding of flight deck automation.

The prevalence of the automation issue “Mode transitions may be uncommanded by pilot” also seemed to differ between these populations; it ranked 12th in frequency in the regional airline analysis and 4th in frequency in the major transport analysis (see Table 3). This issue was used to identify situations in which automation may produce surprising behavior by changing modes without pilot commands. This type of incident may occur while the aircraft is under the guidance of the FMS. As identified in the previous section,

Table 3. Flight Deck Automation Issues in Incident Reports.

<table>
<thead>
<tr>
<th>Automation Issue Statement</th>
<th>Regional Rank</th>
<th>Regional % reports (# incidents)</th>
<th>Transport Rank</th>
<th>Transport % reports (# incidents)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilots may be overconfident in automation</td>
<td>1</td>
<td>46.43% (39)</td>
<td>1</td>
<td>18.31% (52)</td>
</tr>
<tr>
<td>Automation behavior may be unexpected and unexplained</td>
<td>2</td>
<td>40.48% (34)</td>
<td>2</td>
<td>16.20% (46)</td>
</tr>
<tr>
<td>Understanding of automation may be inadequate</td>
<td>3</td>
<td>14.29% (12)</td>
<td>20</td>
<td>1.06% (3)</td>
</tr>
<tr>
<td>Displays (visual and aural) may be poorly designed</td>
<td>4</td>
<td>11.90% (10)</td>
<td>8</td>
<td>3.17% (9)</td>
</tr>
<tr>
<td>Automation may demand attention</td>
<td>5</td>
<td>8.33% (7)</td>
<td>3</td>
<td>10.21% (29)</td>
</tr>
<tr>
<td>False alarms may be frequent</td>
<td>5</td>
<td>8.33% (7)</td>
<td>13</td>
<td>1.76% (5)</td>
</tr>
<tr>
<td>Controls of automation may be poorly designed</td>
<td>7</td>
<td>7.14% (6)</td>
<td>7</td>
<td>3.87% (11)</td>
</tr>
<tr>
<td>Failure assessment may be difficult</td>
<td>8</td>
<td>5.95% (5)</td>
<td>6</td>
<td>4.58% (13)</td>
</tr>
<tr>
<td>Pilots may over-rely on automation</td>
<td>9</td>
<td>4.76% (4)</td>
<td>9</td>
<td>2.46% (7)</td>
</tr>
<tr>
<td>Database may be erroneous or incomplete</td>
<td>9</td>
<td>4.76% (4)</td>
<td>4</td>
<td>4.93% (14)</td>
</tr>
<tr>
<td>Insufficient information may be displayed</td>
<td>11</td>
<td>3.57% (3)</td>
<td>25</td>
<td>0.35% (1)</td>
</tr>
<tr>
<td>Automation may lack reasonable functionality</td>
<td>12</td>
<td>2.38% (2)</td>
<td>18</td>
<td>1.41% (4)</td>
</tr>
<tr>
<td>Failure recovery may be difficult</td>
<td>12</td>
<td>2.38% (2)</td>
<td>24</td>
<td>0.70% (2)</td>
</tr>
<tr>
<td>Mode transitions may be uncommanded by pilot</td>
<td>12</td>
<td>2.38% (2)</td>
<td>4</td>
<td>4.93% (14)</td>
</tr>
<tr>
<td>Pilots may under-rely on automation</td>
<td>12</td>
<td>2.38% (2)</td>
<td>24</td>
<td>0.70% (2)</td>
</tr>
<tr>
<td>Data entry and programming may be difficult and time consuming</td>
<td>12</td>
<td>2.38% (2)</td>
<td>18</td>
<td>1.41% (4)</td>
</tr>
<tr>
<td>Standardization may be lacking</td>
<td>12</td>
<td>2.38% (2)</td>
<td>n/a</td>
<td>0.00% (0)</td>
</tr>
</tbody>
</table>
a much greater percentage of major transport aircraft types are equipped with an FMS compared to regional airlines. The difference in the percentage of aircraft with an FMS in each of these populations may, in part, account for the variation in frequency for this issue.

SUMMARY

As we have tried to understand the regional airlines and their current use of automation, it has become clear that there are enough differences between the regional airlines and the transport airlines that research focusing specifically on their needs is required. We have also discovered how difficult it is to define the needs for addressing automation-related topics without better understanding our definitions of automation and how it affects pilot performance. We developed an expanded pilot general task list and definitions of the three categories of automation to help us accomplish our research goals and found that their use results a better methodology for us to explore automation effects and translate them to the needs of the regional airlines.

ACKNOWLEDGMENTS

This work was supported in part by subcontract C-794014 from San Francisco State University under a NASA grant to SFSU. Dr. Kathy Mosier is the Principal Investigator for the NASA grant and has contributed to the work and the development of these concepts. Dr. Judith Orasanu is the NASA contract monitor for the grant.

REFERENCES

9. AUTOMATION AND AGING: ISSUES AND CONSIDERATIONS

Mustapha Mouloua, Janan Al-Awar Smither, Dennis A. Vincenzi and Laura Smith

INTRODUCTION

With the advent of modern digital technology and the miniaturization of computer technology, it is now possible to integrate highly sophisticated electronic components into systems where it was not possible to do so ten years ago. Computers are ever vigilant, do not tire or fatigue after extended periods of work, and do not require a break in order to return to a previously higher state of vigilance.

Automation can be defined as the execution by a machine agent (usually a computer) of a function that was previously carried out by a human (Parasuraman & Riley, 1997). The development and application of highly reliable automated systems in today’s world has changed the role of the human from an active system operator to one of a passive system monitor, a role for which humans are not well suited (Parasuraman, 1987).

Monitoring of highly automated systems is a major concern for human performance efficiency and system safety in a wide variety of human-machine systems (Mouloua & Parasuraman, 1994; Parasuraman & Mouloua, 1996, Mouloua & Koonce, 1997; Scerbo & Mouloua, 1999).

In the present chapter, we will first review some of the pertinent issues concerning advanced automation technology across a wide variety of systems. Then, we will address the nature of aging as it relates to advanced technological systems. Lastly, we will draw a conclusion on current and future trends of automation technology as they apply to the elderly population.

**BENEFITS OF AUTOMATION**

There are numerous benefits to automation across a variety of applications (Gaba, 1994; Rouse, 1991, Woods, 1994; Thompson, 1994; Mouloua & Parasuraman, 1994; Parasuraman & Mouloua, 1996; Mouloua & Koonce, 1997, Scerbo & Mouloua, 1999; Endsley & Khaber, 2001). Reduction of human error is a major benefit and point of justification for introduction and use of automation into a system. When automation is applied for reasons of safety, it is often because a particular incident or accident identifies cases where human error was seen to be a major contributing factor (Parasuraman, 1997). Designers of automated equipment attempt to remove the source of error by automating functions carried out by the human operator (Parasuraman, 1997). Controlled Flight Into Terrain is one type of human error related accident that has been reduced greatly by the introduction of Ground Proximity Warning Systems (GPWS). The GPWS is an automatic monitoring and alarm system that warns the pilot if the plane is flying too close to the ground or is approaching the ground too rapidly. Since positive events are not typically recorded, it is not possible to document how many crashes have been avoided because of the presence of the GPWS in the modern cockpit (Wickens, 1992). Statistics on the reduced ground collision accident rate since its mandatory incorporation in 1974, however, suggest that the presence of Ground Proximity Warning Systems in the cockpit has increased safety (Wickens, 1992).

Human error, however, can never be totally eliminated. By replacing the human operator with automated equipment in an attempt to eliminate human error, another potential form of human error emerges. The perfection and reliability of the automated equipment is limited by the hardware and software capabilities that are designed and built by humans. To the extent that a system is made less vulnerable to operator error through the introduction of automation, it is made more vulnerable to designer error (Parasuraman, 1997).

Overall, it appears that the introduction of automation into complex systems has greatly increased the level of safety and reduced human error. An examination of air carrier accidents suggests that newer, more highly automated aircraft have had substantially fewer accidents than earlier aircraft (Billings, 1997).
The reliability of automated systems has also improved greatly over the past two decades. Advances in technology and improved displays (glass cockpits and head-up displays) have played a major role in the increased reliability associated with automated systems. One factor affecting automation usage is the operator’s trust in the automation. Unreliable automation is unlikely to be trusted by the operator, and therefore will not be used. Parasuraman, Molloy and Singh (1993) found that even after the simulated catastrophic failure of a highly reliable automated engine-monitoring system, participants continued to rely on the automation for a period of time.

When the overall level of automation reliability is perceived as relatively high, operators may come to rely on the automation, so that occasional failures of the automation do not substantially reduce trust or reliance on automation unless the failures are sustained for a period of time (Parasuraman, 1997). When placed in a situation where the automation is highly reliable and the workload is high, monitoring of automated equipment will be poor.

Complacency has been documented in numerous real life situations where monitoring of automated machinery is boring and uneventful due to the high reliability of the automation. In spite of some of the potentially negative aspects of highly reliable systems, the human operator’s trust in the automated system is critical to the acceptance, operation and success of the human-machine system as a whole (Wickens, 1992).

Also, economy and comfort have been improved through more sophisticated monitoring of both environmental and system parameters. Many airlines require maximum use of automation during flight operations. Use of autopilots, autothrottle, and autoland technologies result in much more efficient operation, and smoother power transitions through use of pre-programmed algorithms, that result in significant cost savings over time. Automated control of an aircraft can produce much smoother, more precise operation of complex systems resulting in greater comfort, greater fuel efficiency, and greater overall efficiency of operation. Comfort has also been improved by the ability of newer aircraft to fly at higher altitudes above most weather (Billings, 1997).

**THE ROLE OF THE HUMAN IN AUTOMATION**

Human monitoring of automated systems for malfunctions in the real world can often be poor as a result of low frequency of occurrences of automation failures when dealing with reliable automated systems. Empirical evidence in a controlled setting has been scant until a study by Parasuraman, Molloy, and Singh was conducted in 1993. In this study non-pilot subjects performed a tracking task and a fuel management task manually over four 30-minute
sessions. At the same time, an automated engine-status task had to be monitored for occasional automation “failures” (engine malfunctions not detected by the automation systems). In another condition, the engine-status task was also performed manually. Subjects detected over 75% of malfunctions on the engine-status task when they did the task manually, while simultaneously carrying out tracking or fuel management. However, when the engine-status task was under automation control, there was a marked reduction in operator detection rate of system malfunctions (i.e. automation failures) (mean = 32%). This substantial reduction in failure detection sensitivity was apparent after about 20 minutes spent under automation control. In a separate experiment conducted during the same study, it was shown that monitoring of automation was poor only under multi-task and not single-task conditions (i.e. when only the engine-status task had to be performed). In a follow-up study, experienced pilots were also found to show similar performance trends (Parasuraman, Mouloua & Molloy, 1994). The findings clearly indicate that highly reliable automated systems are not monitored well by human operators when they are engaged in other manual tasks, probably due to the additional demands placed on existing attentional resources (Mouloua & Parasuraman, 1995) and over-reliance or over-trust in automation (Mosier, Skitka & Korte, 1994; Riley, 1994).

Although poor monitoring of automation appears to be a general finding under multi-task performance conditions, individuals may differ in the extent to which they exhibit this phenomenon. Parasuraman, Mouloua and Singh (1993) found some evidence to suggest that individuals with high self-reported levels of “energetic arousal” showed better monitoring of automated systems (at least temporarily). Older adults may have lower levels of energetic arousal than younger adults and may suffer from vigilance loss, an aspect of monitoring found in highly reliable automated systems (Giambra & Quilter, 1988; Mouloua & Parasuraman, 1995). Thus, older adults may be more vulnerable than younger adults to automation-related reductions in monitoring efficiency.

Researchers have also looked at monitoring of automation failures by young and old adults in dual-task situations (Tsang, 1992; Tsang, Shaner & Schnopp-Wyatt, 1995; Tsang & Voss, 1996; Wickens, Braune & Stokes, 1987). The results of these studies indicated no significant differences in performance between young and old adults in a dual-task situation. Dual-task studies in which subjects perform two tasks simultaneously, however, have found a greater divided-attention cost in older adults compared to younger adults (Hartley, 1992). Parasuraman et al. (1993) found that monitoring of automation failures was poor in young subjects only when subjects were placed in high workload,
multi-task situations. Hardy et al. (1995) hypothesized that high workload levels increase age differences in attentional performance. It was predicted that older adults in a dual-task situation with an automation routine would show poorer detection of automation failures than younger adults. This hypothesis, however, was not supported by the data collected. No support was found for the prediction that monitoring of automation would be poorer in older adults than in younger adults.

Charles Billings (1997) states that automation has put forth many benefits such as greater economic efficiency, greater safety, improved performance and greater reliability. But as always happens with new technology, automation has brought new costs as well. While some types of failures and errors have been eliminated by automation, new types of failures and errors have been enabled. Some of the new problems are more difficult to prevent because automation is now so capable that it has distanced humans from many aspects of their operations. Humans have, in many cases, been “taken out of the loop.”

Some automated systems today are so complex and so autonomous that the need for feedback to human counterparts (in many cases) is deemed unnecessary. As a result, the human operator is often left in a state of confusion with regards to what the automated system is doing and why. When this situation occurs, the human operator/system monitor will begin to exhibit automation induced complacency, and loss of situation awareness may result. Wiener (1981) defines complacency as “a psychological state characterized by a low index of suspicion.” Endsley (1994) formally defines situation awareness as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.”

Some of the problems arising from the automated system’s inattention to human counterpart include performance decrements, loss of situation awareness, development of poor mental models, and increased mental workload as a result of this lack of understanding. As long as the highly reliable, automated system does not fail, these human-centered inadequacies do not pose a major problem. The system will react as programmed, respond as needed and perform its duties and responsibilities in an acceptable and efficient manner.

Advanced automated systems act at a fairly high level of authority and autonomy. Input from a variety of sources other than the pilot (e.g. sensors of the airplane environment, designer instructions) is capable of triggering automated behavior (Sarter, 1997). This high level of authority and autonomy of modern cockpit automation has played a role in recent incidents and accidents (Sarter, 1997).
If and when the automated system fails, whether it is part of a nuclear power plant or part of a Boeing 767, the system may need to revert to manual human control. When this situation occurs, if the operator is not ready to deal with such a situation, numerous human-machine interaction problems could potentially occur. As long as the human is ultimately responsible for the system and the consequences when the system fails, the human must be capable of taking control whenever he or she deems necessary. In order to accomplish this goal, the human must be kept informed of the system and the environment at all times. The automated system must be a “team player”. Automated systems that are strong, silent, clumsy, and difficult to direct are not team players (Woods, 1996).

**COSTS OF AUTOMATION**

When automation first started to become an integral component of human-machine system, in some ways, it was envisioned as a cure all for problems in aviation. The concept of automating as much as possible was considered appropriate with the justification for and expected benefits of automation being a reduction in pilot workload and increased safety through reduced human error. Although many of these goals have been achieved, the design and behavior of automated systems has interacted with the human component of the system in unanticipated ways to create other unforeseen problems and consequences. Some of the problems associated with automation are increasing system complexity, opacity or lack of understanding of the system and its functioning, increased training requirements, increased mental workload, overtrust and mistrust of the automated system leading to automation induced complacency, out of the loop unfamiliarity and loss of skill, and loss of situation awareness.

*System Complexity*

Aircraft automation today is very capable, increasingly flexible and very complex (Billings, 1997). The highly automated glass cockpit offers many advantages to pilots such as automating fuel management procedures which can save fuel and increase efficiency, and automating navigation functions which frees the pilot from having to perform routine psychomotor tasks (Kantowitz & Campbell, 1996). In a more general sense, flightdeck automation can reduce the number of human errors by providing predictable, programmable, and reliable systems as well as reducing the involvement of people in the operation of the system (Kantowitz & Campbell, 1996). From an engineering perspective, flightdeck automation is possible to a very high degree and should be implemented at every possible opportunity.
This view does not, however, take into account the human component of the human-machine system. From a human factors perspective, automation brings with it many problems that manifest themselves when automated systems fail and the system must revert to manual control, or when the human is taken out of the automated environment and is forced to interface with older, non-automated flightdecks. For example, increased monitoring requirements (Lerner, 1983), over reliance on the system (Danaher, 1980), and proliferation of flightdeck components (Wickens, 1992) have all emerged as unintended consequences of automation that can increase pilot workload (Kantowitz & Campbell, 1996).

**Opacity: Lack of Understanding of Automation**

Opacity can be defined as the degree to which an automated system informs the human component of the human-machine system as to what it is doing, why it is doing what it is doing, and what it intends to do next. The three questions which paraphrase the most frequent responses of pilots to automation surprises are; “What is it doing?”, “Why is it doing that?”, and “What is it going to do next?” (Wiener, 1989).

These three questions bring to light two important facets of automated systems that could potentially lead to serious problems when these systems perform in a less than optimal manner. The first problem is that a highly reliable automated environment that does not communicate well with its human counterparts results in a poor mental model of the automated system. This can be due to the complexity of the automated system or to inadequate training or both (Billings, 1997). Another problem is that the automation does not bother to inform the pilot of why it is doing what it is doing and does not explain what it intends to do before it does it. Completely autonomous systems behave in this fashion because they have been designed to do so. The engineers and software programmers who have designed the equipment and programs under which the automation will operate, often fail to consider that a human will be involved in controlling the system. Evidently, the system designers have deemed it unnecessary to inform the pilot of what the system is doing, why it is doing it and what the system intends to do next.

For a pilot to become an effective, integral part of a system a clear mental model of its functioning must be developed, and a clear understanding of the behavior of the system must be achieved. Automation opacity may be deliberate: one sure way to keep the pilot from intervening in a process is to deny him or her the information necessary to permit intervention (Billings, 1997). The effect of a system having a low degree of opacity will result in a
poor mental model being developed, frustration and distrust being fostered, and quite possibly a partial or complete loss of situation awareness.

On the other hand, opacity at some level is required in order to avoid overwhelming the pilot with information and data. The capabilities of the computer and its screens have made it possible for designers to overwhelm pilots with information and data (Billings, 1997). This point accentuates the need to develop displays that do not overwhelm and inundate the pilot with superfluous information. Displays must provide the proper, relevant information that is needed for the pilot to obtain a proper assessment of the situation at hand.

Training

Training requirements are increased by automation with the potential for commensurate changes in pilot workload (Kantowitz & Campbell, 1996). Automation introduces unique training requirements because pilots must become proficient in both manual and automatic flight modes (Kantowitz & Campbell, 1996). In the early 1960s, major airlines such as Trans World Airlines and United Airlines began revising their pilot training programs from a “teaching the pilot how to build the airplane” approach to one that centered on teaching the pilots how to operate the aircraft without burdening them with more systems knowledge than they “needed to know” (Billings, 1997). The major advantage of this was a significant reduction in training cost to the company without any apparent reduction in training effectiveness.

The introduction and complexity of advanced automation has given rise to some serious questions about this approach to training. As indicated previously, pilots must have an adequate mental model of the behavior of the equipment they are flying. The operation of automated equipment without the understanding of “what the system is doing and why it is doing it” has had the profound effect of increasing the mental workload of the pilot.

Pilots must now be trained on and become proficient with manual flight modes, automated flight modes and the transition between manual-to-automated-to-manual flight modes as well as unique requirements for inter-crew communications (Kantowitz & Campbell, 1996). Furthermore, flightdeck automation is not binary, that is, either on or off. There are several modes or levels associated with automated systems, all of which must be learned and remembered by pilots (Kantowitz & Campbell, 1996). The human factors community is well aware of problems attributed to automation in the modern commercial aircraft. Researchers have identified “lack of mode awareness” as one of the most critical of these problems (Andre & Degani, 1997).
Modes represent the different behaviors, or functions, of a given system. The more functions a system has, the more potential modes will be available (Andre & Degani, 1997). Mode awareness refers to the operator’s knowledge and understanding of the current and future status and behavior of the system, given a particular mode. Mode errors have been directly linked to several recent accidents involving modern transport aircraft (Andre & Degani, 1997; Hughes & Dornheim, 1995). Pilots must become familiar with both the system capabilities and operator responsibilities at each level of automation. Flying without automation provides the pilot with manual flight experience. Conversely, flying with automation deprives the pilot of practice in the manual mode, which may induce a loss of proficiency that requires additional training time to return to the previous level of proficiency (Kantowitz & Campbell, 1996). As flightdeck automation becomes more capable and is able to do more, the systems will become increasingly more complex and complete comprehension will be much more difficult.

*Increased Mental Workload*

Pilot workload is defined as an intervening variable, similar to attention, that modulates or indexes the relationship between the demands of the environment and the capacity of the operator (Kantowitz, 1994). In other words, mental workload is the ratio of the amount of resources demanded by a task or multiple tasks and the amount of resources available within an individual. Workload itself is a construct that must be studied and evaluated indirectly; it cannot be directly evaluated or observed. It is a conceptual, multifaceted construct that must be inferred from changes in observable data (Kantowitz & Campbell, 1996). The complexity of workload has resulted in its measurement by many methods (Moray, 1982). Pilot workload is measured by evaluating subjective ratings, objective performance on primary, secondary and other incidental tasks, and physiological measures.

Mental workload is subjective and highly individualistic. The amount of mental workload experienced by one individual may be very high in one particular situation whereas a different individual placed in a similar situation may experience average or low levels of workload. Much of the individual’s workload sensation will depend on a great many factors such as amount of resources available, difficulty and complexity of the task or situation, and the individual’s familiarity or unfamiliarity with the task or situation, and how capable the individual is of dealing with multiple tasks.

Human factors research has demonstrated that many of the errors and accidents that occur during task performance are associated with levels and types
of operator workload (Kantowitz & Campbell, 1996). Factors that can influence pilot workload have been grouped according to how they affect workload. In general, increases in factors such as pilot skill and experience, feedback and system reliability will serve to decrease workload whereas increases in factors such as task demands, environmental demands, and pilot fatigue will serve to increase workload. The relationship between factors such as pilot skill, feedback, and system reliability and factors such as task demands, environmental demands, and pilot fatigue is very dynamic in nature. The resultant emergent capability of the pilot produced by the combination of pilot skill, feedback, and system reliability is then summed together with the resultant situational demands to produce an overall state of subjective mental workload that the pilot experiences. Pilot skill, feedback, and system reliability can be considered the pilot’s available resources whereas task demands, environmental demands, and pilot fatigue can be considered the pilots situational demands. The proper combination of available resources and situational demands produces an overall pilot workload that will result in optimal performance.

Although increases in automation were theoretically believed to lower a pilot’s workload, research has indicated that automation has had the reverse effect; that is automation increases mental workload (Mouloua, Parasuraman & Molloy, 1993; Parasuraman, Mouloua, Molloy & Hilburn, 1992). The reason for this is that although automation does decrease a pilot’s workload in some areas, it also causes increases in workload in other areas. A major aspect of automation is that the pilot’s role is being changed from that of an active participant to one of a passive monitor. Although the automated system may relieve the pilot of manually flying the aircraft, the pilot must now monitor the normal system parameters and also monitor the automated system to ascertain that the automated system is functioning properly. This increased monitoring requirement has had a profound effect on the human operator, and increases in automation have led to a redefinition of the operator (Kantowitz & Campbell, 1996).

One reason for introducing automation in complex systems was to reduce the chance of human error by reducing the operator’s high mental workload. This is not always the case. The view that automation does not necessarily reduce mental workload was first pointed out over twenty years ago (Edwards, 1976). Increase in automation causes an increase in vigilance and monitoring requirements, leading to increases in mental workload.

Over Trust and Mistrust

The human operator’s trust in an automated device is critical to the smooth functioning of the human-machine system. However, this trust must be tempered
with a touch of reality. Automated systems have great capabilities and have proven themselves to be safe, reliable and cost efficient, however, the operator of the system must keep in mind that automated systems are no more than computers and software that are capable of failure.

The pilot’s level of trust in the system can have an important effect on workload and subsequent performance. Research has indicated that pilots given highly reliable automated devices will come, over time, to rely on the assistance they provide. However, operating highly reliable automated equipment over long periods of time can create very high levels of trust in the system, and consequently, the pilot may eventually become over-reliant on the system. In some cases, pilots have continued to use automation even when they had every reason to mistrust it. It does little good to remind human operators that automation is not always reliable or trustworthy when their own experience tells them it can be trusted to perform correctly over long periods of time (Billings, 1997).

Lack of trust of an automated system is also detrimental to performance due to the fact that the automated system that is not trusted will more than likely not be used. Levels of trust in automated systems that are inappropriately low can lead to mental workload levels that are too high, whereas levels of trust that are inappropriately high can lead to mental workload levels that are too low (Kantowitz & Campbell, 1996).

Out of the loop Unfamiliarity

As humans practice a task, they become more familiar with the all factors involved in any particular function. As the individual becomes more familiar, practiced and comfortable with a task, the manual, perceptual and cognitive processes involved in performance of the task shift from requiring great amounts of attention and cognitive resources to one where the task becomes automated, and much of the processing and performance of the actual task is done in the background. Information from the environment is absorbed very quickly and “chunked” into blocks of information. Only the necessary information needed to perform the task efficiently is extracted from the environment for use.

A good example of a task, which initially requires huge amounts of cognitive resources to perform at an acceptable level, is that of driving an automobile. The actual task of driving an automobile is highly dynamic in nature in that the environment is constantly changing. One must be aware of not only the external factors that may affect safety and performance, but internal factors as well. One must constantly be monitoring the actions of other drivers as well as the operational parameters of oneself and the vehicle that the
individual is operating. In addition to monitoring the internal and external parameters of the environment, the driver is also operating the vehicle in terms of speed and directional control. Information taken from the environment is processed and translated into performance while operating the vehicle.

When automation is introduced into a system, two things happen. First, the operator is transformed from an active operator to a passive monitor. In doing so, the increased monitoring load placed on monitors of automated systems may not free up an operators attentional resources to perform other duties as is frequently the intended goal of automation (Kantowitz & Campbell, 1996). Second, in changing the individual’s role from an active participant and operator to one of a passive monitor, the individual is effectively removed from the active control loop and loss of familiarity with the key system elements and processes for which he or she is responsible will result (Kantowitz & Campbell, 1996). In effectively taking the operator “out of the loop”, the operator’s skills may erode to the point that he or she is no longer an effective, efficient part of the system. In high workload, highly dynamic environments such as driving an automobile, flying an aircraft or operating a nuclear power plant, efficiency of operation and familiarity of the system are essential to effective and safe operation of the system. This effect seems to be exacerbated in emergency situations where time is critical, danger is imminent and system failure results.

In a number of experiments investigating participatory modes in automated systems, active controllers had exhibited faster response times and more accurate failure detection performance than passive monitors (Wickens & Kessel, 1981). Thus, as the distance between the operator and the system under control increases, workload can be increased if the operator is suddenly required to jump back into the active control loop and directly control the system (Kantowitz & Campbell, 1996). Long term participation as a system monitor rather than as an active controller can also lead to reductions in baseline skill levels and reduced decision making abilities, particularly for highly automated system functions (Kantowitz & Campbell, 1996).

One proposed method of combating this problem lies in proper implementation of adaptive task allocation or adaptive automation. Adaptive automation refers to a system capable of dynamic, workload-triggered reallocations of task responsibility between human and machine (Hilburn, Jorna, Byrne & Parasuraman, 1997; Parasuraman, Mouloua & Molloy, 1996). The traditional approach to allocation of functions in a human-machine system was originally determined by a dichotomous relationship which gave full control of a task either to the human or to the machine. In today’s technological society it is debatable whether some of the functions that need to be performed are better
executed by a human or by a machine. Automation provides a number of benefits, but also has some serious costs associated with its implementation. According to proponents of adaptive systems, the benefits of automation can be maximized and the costs minimized if tasks are allocated to automated subsystems or to the human operator in an adaptive, flexible manner rather than in an all-or-non fashion (Rouse, 1988). An ideal situation would be one where the operator could switch the control of a task from manual to automated when workload conditions are high. Once the operator’s workload was reduced, the operator could then continue to perform the task in manual mode, thereby maintaining familiarity with the system and preserving the operator’s cognitive ability and baseline skill level.

**Loss of Situation Awareness**

In many instances, operators of highly reliable automated systems fail to recognize when automated systems are malfunctioning due to the fact that they have been taken “out of the loop.” Technological advances have been associated with a number of problems including loss of situation awareness. Situation Awareness (SA), is loosely defined as an operator’s internal model of the surrounding. Operators of dynamic systems must be aware of the current status of their particular system and their environment in order to be able to predict the future status of their specific situation. This will allow the operator to decide on the best and most appropriate course of action possible in any given situation. Many of the elements discussed thus far ultimately culminate in “loss of situation awareness.”

The changing of roles of the human from an active participant in the system to one of a passive monitor is a major problem. The human operator becomes a passive observer of an automated system rather than an active processor of information. Evidence suggests that the very act of becoming passive in the processing of information may be inferior to active processing (Cowan, 1988). This factor could make a dynamic update of system information and integration of that information in active working memory more difficult (Endsley, 1994).

Complacency is another major factor associated with a lack of vigilance in monitoring automated systems and development of low situation awareness. Development of a complacent attitude toward an automated system is a direct result of the amount of trust and confidence an individual places in the automated system. Too high a degree of trust will lead to complacency and over trust in a system. Operators may elect to neglect the automated system
and the system parameters overseen by the automation in favor of other tasks through a shifting of attention (Parasuraman, Mouloua & Molloy, 1994).

Lack of understanding of the automated system is another major factor contributing to a decrease in situation awareness. Some systems are so complex that many operators have difficulty understanding what the systems are doing and why, even when the systems are performing as intended. An explanation of this phenomenon may lie in the inherent complexity associated with many of these automated systems, poor interface design, inadequate training or a combination of these factors (Endsley, 1994).

The traditional form of automation that places humans in the role of monitor has been shown to negatively impact situation awareness and, thus, the skills and abilities of the system operator. Reduction in situation awareness can be attributed to a number of factors including, but not limited to, poor training with regards to the automated system, poor feedback by the system, poor interface design, development of complacent attitudes, over-reliance on highly reliable automated systems and increased workload when monitoring automated systems.

**HOW DO OLDER ADULTS COPE WITH AUTOMATION?**

A major concern associated with the aging population of today is how age affects an individual's ability to learn and deal with new and unfamiliar situations. Advances in automation technology have created many improvements in society, but the question remains as to how the elderly population will respond to new technologies. If monitoring of automated systems presents challenges to younger individuals who do not suffer from the deleterious effects of aging, how will older individuals interact with automated systems and increased demands for efficient monitoring skills? Questions such as these need to be addressed if older individuals will be required to interact with new technologies.

It is estimated that by the year 2025, approximately 20% of the U.S. population will be 65 years of age or older (Harbin, 1991). Certain areas relevant to aging, mostly physiological in nature, have been researched quite thoroughly. The effects of these age related changes on the individual have been well documented. It is well known, for example, how the aged eye can affect the interaction between the individual and the real world environment. It is also well documented as to how physical deterioration due to age can affect a person in the real world. One aspect of aging that is not well documented or understood is how people will react, adapt and develop, with respect to
cognitive ability in this new high tech automated world that is taking shape around us. Along with the development and advancement of life prolonging medical technologies, advanced electronics, and the spread of automation across several human-machine systems (e.g. transportation, nuclear, medical, etc.), individuals are kept active in the work place for longer periods of time. Thus, those who will be working longer must develop new skills to adapt to their fast changing work environment.

As applications of automated technology continue to increase, the demand for highly developed monitoring skills will also increase. Although a lifetime of working with advanced technologies may promote development of high levels of expertise that may allow an individual to perform better and function more efficiently in these new environments, expertise may not necessarily ameliorate all the effects of aging that impact older adults’ performance with automated systems. As mentioned earlier, research indicates that older adults may have lower levels of energetic arousal than younger adults, leading to reduced vigilance in simple tasks (Giambra & Quilter, 1988). They may also exhibit poor performance in more complex vigilance tasks due to age-related reduction in attentional capacity (Mouloua & Parasuraman, 1995). These factors suggest that older adults may be more vulnerable than younger adults to exhibit automation-related complacency.

Many cognitive functions, such as divided attention and working memory, are known to decline in efficiency with adult aging. Other functions, such as general and system-specific knowledge, may improve (Birren & Schaie, 1996; Hawkins, Kramer & Capaldi, 1992; Salthouse, 1985). Previous studies on the effects of aging on sustained attention or vigilance have shown that differences between young and old adults are dependent on the nature of the vigilance task performed (Davies & Parasuraman, 1982). The target detection rate in low event-rate vigilance tasks appears insensitive to the effects of adult age (Giambra & Quilter, 1988). However, recent evidence suggests that adult age differences in vigilance performance are magnified in high-event rate tasks and when taskload is increased (Mouloua & Parasuraman, 1995). In these tasks older adults had lower detection rates and lower levels of perceptual sensitivity than young adults. A high stimulus event rate has been shown to lead to decrements in perceptual sensitivity over time, presumably because of the need to maintain controlled or effortful processing at a high rate for long periods of time (Parasuraman, 1985).

Many studies have compared the performance of young and old adults in single-task and dual-task situations (Tsang, 1992; Tsang, Shaner & Schnopp-Wyatt, 1995; Tsang & Voss, 1996; Wickens, Braune & Stokes, 1987; Kramer, Hahn & Gopher, 1999). These studies have generally found that even
when single-task performance is equated for difficulty between young and old subjects, dual-task performance is poorer in older adults. This age difference has been attributed to an age-related divided attention cost, or to reduced ability with age to allocate processing resources efficiently to multiple tasks (Hartley, 1992).

Examination of the influence of aging on automation complacency is important for several reasons. First, the aging of the general population applies to the general work force as well as to workers in high technology industries. Further, the increased application of automation in industry, particularly in aviation, merits further examination of the impact of automation on human performance in older workers. An additional motivating factor in aviation in the U.S. (and in most countries worldwide, with some variation), involves the mandatory retirement of all commercial airline pilots by age 60. This so-called Age 60 rule was instituted several decades ago on the basis of then-current standards on physical health. The rule has led to a controversial debate in the last few years and scientists and pilots have often testified before the Congress to dispute the rule. As pointed out by Hardy and Parasuraman (1997) in a recent review of the cognitive aging literature relevant to the Age 60 rule, although much research has been carried out on aging and human performance (Salthouse, 1985), there is a dearth of empirical data on the performance of older adults on complex tasks of the type encountered in aviation and other human-machine systems.

Mouloua, Vincenzi, Hardy, and Parasuraman (under review) conducted a series of studies to investigate the performance of older adults in monitoring automated systems. A goal of their first experiment was to investigate baseline levels of performance. Results showed that older adults could perform the system monitoring task as well as young adults. Furthermore, the results also replicated the previous findings of Parasuraman et al. (1993) that indicate monitoring for system malfunctions is equally efficient under manual and automated conditions when it is the only task being performed, i.e. automation complacency does not occur under single-task conditions.

As predicted, operator detection rate of engine malfunctions in the dual-task flight-simulation was significantly lower when the task was automated than when it was performed manually. This finding is consistent with previous results of Parasuraman et al. (1993) in which subjects performed three flight-related tasks, but extends that study by showing that automation complacency can also occur under dual-task conditions.

No support was found for the prediction that monitoring of automation would be poorer in older adults than in younger adults. A greater effect of automation complacency in older adults was predicted because it has been
shown that age differences in attentional performance are exacerbated in high taskload situations (Hartley, 1992; Mouloua & Parasuraman, 1995). Both groups showed equivalent detection performance in the manual condition. However, while there was a decline in detection rate in the automated condition, there were no significant group differences in performance in this condition. Thus, while both the young and old groups exhibited the automation complacency effect, the extent of the effect did not differ between groups.

Although no significant age effects were obtained in this experiment, older adults typically had lower mean levels of detection performance in the automated condition, particularly in the later blocks. This suggests that an age-related increase in automation complacency may exist, but that it is small under dual-task conditions. One possible explanation for the finding that monitoring of automation was equally inefficient in young and older adults may be that the level of processing capacity demands for each age group might be a factor. If capacity demand is an issue, then it would be fruitful to examine the performance of young and old adults as task load is increased further. This line of reasoning prompted the researchers to carry out a similar experiment in young and old adults under multi-task conditions. This second experiment replicated and extended the previous results of Parasuraman, Molloy, and Singh (1993) on the conditions that lead to automation-induced complacency. The results indicated that age differences in system monitoring performance are obtained only if task demands increase to a point where resources available to perform a task are exceeded. While older adults showed a significant decrease in monitoring performance in the multi-task condition, single- and dual-task performance were unaffected by adult aging. In the multi-task condition, detection rate increased for the younger group as a function of time on task, whereas the older group showed a decline over time. Thus, the hypothesis that automation-induced complacency would be greater in older adults than in younger adults was supported in the multi-task condition. It can be concluded that the performance cost of automation-induced complacency is more pronounced in the older group than in the younger age group only under high task load conditions.

One possible explanation for the differences between age groups on detection of automation failures is that younger participants have better attention allocation control, and are more capable of initiating and sustaining effortful processing. Attention allocation control refers to the extent to which one can direct one’s attention to different tasks according to other task demands (Tsang & Voss, 1996). Older participants may have less attention allocation control than younger participants, and may not be able to maintain adequate attention allocation control in the presence of high workload situations involving multiple
tasks being performed simultaneously. Older participants may not be capable of maintaining constant levels of attention allocation unless a task requires active participation as in the case of tracking and resource management. As a result, performance in the system monitoring task suffers. Participants who do not allocate sufficient attentional resources to the system monitoring task may be looking but not seeing (Molloy, 1996).

The researchers also hypothesized that the subjective workload experienced by the older group would be significantly greater than the subjective workload experienced by the younger group. This hypothesis was not supported. Subjective workload did not appear to vary as a function of age. Both the older and younger groups experienced comparably high levels of subjective workload. The fact that subjective workload expressed by both age groups in the dual task condition was not significantly different from the subjective workload expressed by both age groups in the multi task condition indicates that subjective workload does not necessarily increase as task load increases, even though performance on the specified tasks changes significantly. In the single task condition, the subjective workload expressed by the participant was relatively low, with the overall mean averaged across all four sessions equal to 51.0. As the task load was increased by adding an additional task to perform in the dual task condition, overall workload across sessions increases to 64.8. However, as the task load was increased even further by adding a third task in the multi task condition, the overall subjective workload expressed by the participant increased only slightly to 66.1. This apparent “saturation” effect suggests a dissociation between the effects of task load on subjective measures and objective performance. Neither younger nor older participants experienced a significant increase in workload from dual to multi – task conditions, yet the older participants performed worse indicating an inability to allocate resources efficiently across all tasks.

Reducing the Performance Gap

The above results suggest that complacency represents one human performance cost of high-level automation. Wickens (1994) has suggested that in addition to complacency, automation that replaces human decision making functions can also lead to loss of situation awareness and skill loss. Collectively he referred to these three costs as reflecting “out-of-the-loop unfamiliarity.” How can such costs be mitigated? One proposed method of maintaining high levels of monitoring performance lies in implementation of adaptive task allocation or adaptive automation. Adaptive automation refers to a system capable of dynamic, workload-triggered reallocations of task responsibility between human
and machine (Hilburn, Jorna, Byrne & Parasuraman, 1997). The traditional approach to allocation of functions in a human-machine system was originally determined by a dichotomous relationship, which gave full control of a task either to the human, or to the machine. In today’s technological society it is debatable whether some of the functions that need to be performed are better executed by a human or by a machine (Fitts, 1951). According to proponents of adaptive systems, the benefits of automation can be maximized and the costs minimized if tasks are allocated to automated subsystems or to the human operator in an adaptive, flexible manner rather than in an all-or-non fashion (Rouse, 1988). An ideal situation would be one where the operator could switch the control of a task from manual to automated when workload conditions are high. Once the operator’s workload was reduced, the operator could then continue to perform the task in manual mode, thereby maintaining familiarity with the system and preserving the operator’s cognitive ability and baseline skill level. The flexibility of an adaptive automated system can also potentially optimize individual differences such as in younger and older operators.

Operator Expertise

One perspective on aging is that with age, multiple decrements, both physical and cognitive, develop as a direct result of the aging process. Another perspective on aging is that aging is a lifelong process of accumulating and organizing knowledge and experience (Morrow, 1996). Research in this area indicates that the organization and accumulation of knowledge and experience seems to offset age related decrements related to cognitive functioning.

A number of problems begin to manifest themselves, showing declines in performance as young as 40 years of age. Sensory functions such as vision and hearing begin to degrade, central nervous systems functioning slows, psychomotor functioning begins to degrade, and declines in perceptual processing and cognitive processing occur (Gerathewohl, 1978a).

Aging, however, is not a unitary process (Braune & Wickens, 1984; Gerathewohl, 1977; Salthouse, 1990). One of the most important variables involved in determinable age changes concerns the individual variance of the age-related functions that affect the validity of the functional age model and its application to the controversy about forced retirement (Gerathewohl, 1978a). Many studies of cognitive aging reveal enormous variability among older adults (Braune & Wickens, 1984; Gerathewohl, 1977, 1978a). There are usually greater differences among a group of older people than there are among the young.
Two questions arise from these views on aging: (1) does experience affect performance in monitoring automated systems, and (2) are the effects of experience robust enough to offset age related declines in performance. Research by Vincenzi and Mouloua (1998) investigated these questions by having both young and old pilots and non-pilots participate in their study. If experience does make a difference in performance, the performance measures recorded for young and old non-pilots should be significantly different from those of young and old pilots. It was expected that performance differences would be found and that experience would offset some age related differences.

A total of 120 male participants, 60 young adults ranging in age from 18 to 30 years (mean age = 24.9 years), and 60 older adults ranging in age from 60 to 75 years (mean age = 66.4 years), participated in this study. Both the young and old group consisted of two separate groups, each containing 30 non-pilots and 30 pilots.

A modified version of the Multi-Attribute Task Battery (MAT) developed by Comstock and Arnegard (1992) was used in this study. The modified version of the MAT battery consisted of two dimensional compensatory tracking, system monitoring, and fuel management tasks. The tracking and fuel management tasks were always performed manually, while the system monitoring task was controlled by an automation routine. The three tasks were displayed in separate windows on the monitor.

The system monitoring task consisted of four vertical gauges with moving pointers and green and red warning lights. Occasional “system malfunctions” occurred at random intervals, indicated by the pointer on one of the four engine gauges going “off limits.” These system malfunctions were normally detected and reset automatically, successful identification and correction of the malfunction being indicated by a red warning light, which was extinguished when the problem was corrected. However, from time to time the automation failed to detect a malfunction. The automation failure rate, which was constant across time, was 7 malfunctions per 10-minute block. Subjects were responsible for detecting automation failures.

A two dimensional compensatory tracking task with joystick control was presented in one window of the display. The subject’s task was to keep the aircraft within the central rectangle by applying the appropriate control inputs in the x- and y- directions.

The resource management task was a simulation of the actions needed to manage the fuel system of the aircraft. Subjects were required to maintain a specific fuel level within both of the main tanks, by selectively activating pumps to keep pace with the fuel consumption in these main tanks.
Performance means and standard deviations were calculated for detection rate, reaction time, resource management RMS error, tracking RMS error, and overall system performance (OSP). With regard to detection rate of automation failures, younger participants in general detected more automation failures (73.12%) than did older participants (50.40%), and pilots performed better on detection of automation failures (69.77%) than non-pilots (53.75%). When the means and standard deviations for each individual group were examined, young pilots detected more automation failures (79.70%), followed by young non-pilots (66.55%), followed by old pilots (59.85%) and lastly by old non-pilots (40.96%). The standard deviations obtained for the young and old pilots were very similar (18.17 and 18.08 for young and old pilots respectively), and the standard deviations obtained for young and old non-pilots were very similar (26.07 and 25.32 for young and old non-pilots respectively). This is an interesting finding because research on aging generally indicates an increase in variability as age increases. In this case, variability is affected by experience as well as age. Both younger and older non-pilots exhibited a much greater degree of variability than younger or older pilots.

Similar performance patterns were found upon examination of reaction time, and resource management. Mean tracking RMS error, however, resulted in younger non-pilots performing the best (208.64), followed by young pilots (258.47), followed by old pilots (262.35), and lastly by old non-pilots (316.76). Younger pilots exhibited the least variability (67.79), young non-pilots and old pilots exhibited similar variability (84.25 and 84.21 for young non-pilots and old pilots, respectively), and older non-pilots exhibited the highest variability (99.63). Young and old pilots exhibited less variability in performance than young or old non-pilots. These findings indicate that qualities such as cognitive performance and time-sharing ability are more stable in pilots than non-pilots in general regardless of age.

Although differences due to age clearly exist, the results of this research appear to indicate that experience does matter with regard to pilot performance. Specifically, experience as a pilot appears to mediate performance and offset some of the deleterious detriments that occur as a result of aging. Individual differences observed in this study clearly indicate that a good portion of the older pilot population was capable of performance that was comparable to or superior to performance exhibited by the younger pilot population.

With the demand for efficient monitoring skills created by development of sophisticated automated human-machine systems on the rise, older individuals will need to develop the skills necessary to remain productive in the later stages of life. The development of new medical technologies that prolong life
and increase its quality will most likely result in an older cohort that is more physically fit and more cognitively capable than in previous generations.

Experience is a potentially powerful moderator of the ability to monitor automation failures. For instance, experience can modulate levels of mental workload. Although mental workload, assessed with a single subjective measure, did not appear to play a role in performance in the above study, it is still a potential mediator of age group differences in monitoring ability in high demand situations. Perhaps age group differences in automation complacency are reduced as a function of expertise. Experience is also related to one’s mental model. Differences in monitoring strategy due to mental models could alter automation complacency in both younger and older system operators. More efficient monitoring strategies should lead to better automation failure detection. On the other hand, experienced operators may also be more comfortable with automated systems, potentially making them more susceptible to automation complacency.

As mentioned previously, since the fastest growing segment of the population is the elderly population, and since the proliferation of automation and technology shows no indication of slowing, further research in this area is warranted. The examination of age-group differences (or similarities) in higher-order skills appears to be a fruitful path of investigation (e.g. Hardy & Parasuraman, 1997). Therefore, to the degree that automation complacency is affected by experience or domain-dependent knowledge, further exploration of age-related differences in such situations will be useful for building models of monitoring performance for younger and older individuals.

REFERENCES


