SUPPORTING COORDINATION IN WIDELY DISTRIBUTED COGNITIVE SYSTEMS: THE ROLE OF CONFLICT TYPE, TIME PRESSURE, DISPLAY DESIGN, AND TRUST

BY

WESLEY ALLAN OLSON

B.S., United States Air Force Academy, 1985
M.S., University of Illinois, 1987

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1999

Urbana, Illinois

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 12.Aug.99	3. REPORT TYPE AND DATES COVERED DISSERTATION

4. TITLE AND SUBTITLE SUPPORTING COORDINATION IN WIDLEY DISTRIBUTED COGNITIVE SYSTEMS: THE ROLE OF CONFLICT TYPE, TIME PRESSURE, DISPLAY DESIGN, AND TRUST	5. FUNDING NUMBERS
6. AUTHOR(S) MAJ OLSON WESLEY A	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF ILLINOIS AT URBANA	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) THE DEPARTMENT OF THE AIR FORCE AFIT/CIA, BLDG 125 2950 P STREET WPAFB OH 45433	10. SPONSORING/MONITORING AGENCY REPORT NUMBER FY99-274
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION AVAILABILITY STATEMENT Unlimited distribution In Accordance With AFI 35-205/AFIT Sup 1	12b. DISTRIBUTION CODE
13. ABSTRACT *(Maximum 200 words)*	
14. SUBJECT TERMS	15. NUMBER OF PAGES 169
	16. PRICE CODE

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

SUPPORTING COORDINATION IN WIDELY DISTRIBUTED COGNITIVE SYSTEMS: THE ROLE OF CONFLICT TYPE, TIME PRESSURE, DISPLAY DESIGN, AND TRUST

Wesley Allan Olson, Ph.D.
Department of Psychology
University of Illinois at Urbana-Champaign, 1999
Dr. Nadine B. Sarter, Advisor

Coordination has been defined as the management of dependencies between the goals, tasks, and resources of various agents (Malone and Crowston, 1990). Recently, effective coordination between human and machine agents has become increasingly important due to increasing levels of system autonomy and authority. The coordination strategy most often preferred by human operators due to a perceived high level of control over machine actions is called management-by-consent (Olson and Sarter, 1999). Under this approach, the machine is not allowed to act unless and until the operator has given explicit consent to proposed goals and actions (Billings, 1997). Since preferences do not necessarily translate into superior system performance, and to contribute to a better understanding of successful human-automation coordination, this study investigated the effects of conflict type, time pressure, operator trust, and display design on an operator's ability to provide informed consent. The context for this research was (a simulation of) the highly automated cockpit of a modern aircraft. 30 professional B-757 pilots flew a set of 8 descent scenarios while responding to a series of air traffic control clearances. Each scenario presented pilots with a different conflict that would arise either from the goals specified in the clearance or from the implementation of the clearance by the automation.

Overall, this study found that pilots were often unable to detect conflicts and thus failed to make informed and accurate decisions about proposed machine actions.

Detection performance was particularly poor for conflicts related to clearance implementation, and, within this category, conflicts were most likely to be missed if the automation did more than expected by the pilot. In addition to conflict type, the factors time pressure, high trust in air traffic control, and low trust in automated systems also contributed to poor detection performance. Based on a model of the cognitive processes involved in conflict detection, these findings are explained by the inability of pilots to generate expectations of system behavior that could guide their pre- and post-consent monitoring, as well as the failure of the automation to provide salient and effective feedback on its goals and intended actions in support of data-driven conflict detection. Possible approaches for improving human-machine coordination through more effective information representation and sharing are discussed.

Acknowledgements

This research was supported by a research grant (96-G-043) from the Federal Aviation Administration. The technical monitors for this grant were Dr. Tom McCloy and Dr. Eleana Edens. My tuition, fees, and salary were paid by the U.S. Air Force under the Air Force Institute of Technology, Civilian Institutions program (AFIT/CIGS).

I would like to thank my committee members, Dr. Chris Wickens, Dr. Chuck Hulin, Dr. Patty Jones, Dr. Art Kramer, and Dr. Nadine Sarter for their support, encouragement, and insights over the course of this project. I would especially like to thank Nadine for her inspiration, enthusiasm, dedication, and guidance throughout my graduate studies. I came to Illinois to work with Nadine, and I leave this institution grateful for the opportunity to collaborate with and learn from one of the best.

I would also like to thank the pilots and management from the major US airline that supported this research. Without the advice, access, and facilities provided by this airline, as well as the participation and professionalism of the pilots, this study would not have been possible. Thanks also to Jonathan Sivier who spent many frustrating, confusing, and ultimately successful hours modifying NASA's Stone Soup Simulator. Finally, I would like to thank Mary Welborn and Karen Ayers for their friendship and assistance during my time at Illinois.

To my parents, Roger and Muriel, thank you for your love, support, and encouragement throughout the years. You laid the ground work for my success in this endeavor.

To my boys Nick, Andrew, and Sam, I hope you will remember this time fondly, and will always want to go visit "Daddy's airplane factory." Finally, and most

importantly, to my best friend and wife, Kelly, thank your for all you did during these last three years. In many ways your job was the hardest - having been down this road before, you felt my pain more than I did. Your support, advice, love, and friend ship made this all possible. I am very blessed to have you for my wife and friend.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Successful teamwork requires coordination in order to avoid resource competition, redundancy, and conflicts between team members. Considerable and successful efforts have been made to encourage and support teamwork among human operators in a variety of domains through programs such as crew resource management. With the ever-increasing capabilities, authority and autonomy of modern technology and automated systems, there is a growing need to expand these efforts to support coordination between human and machine team members (Sarter, Woods, and Billings, 1997).

In the early days of automation, machine roles were limited and the pattern of human-automation interaction constrained by the low level of machine capability. The machine was just a tool, and consequently, there was little serious discussion of human and machines as a coordinated team. The growing power and capability of automated systems have allowed machines to evolve from simple tools that merely extend human capabilities to powerful agents which have the capability to pursue their own goals and plan their own actions in response to changing conditions (Hancock, 1993). Consequently, humans and these highly powerful machine agents need to communicate and coordinate their activities to function as a team. To date, the burden of ensuring a match between human and machine goals and actions rests primarily with the human operator whose role is to supervise machine actions and intervene as necessary to avoid problems or conflicts.

This approach has led to difficulties such as automation surprises (Sarter, Woods, and Billings, 1997) which can be explained by increasing levels of automated system autonomy, authority, complexity and coupling along with relatively low observability. Advanced automated systems are capable of initiating long action sequences with little or no operator input (high autonomy) and, in some cases, of overriding operator commands (high authority) (Woods, 1996). Growing system autonomy increases the attentional demands for the operator who must track and anticipate machine actions. It also requires that human and machine actions are coordinated in advance because of the potential time lag between human input and machine action. High levels of system authority can create coordination difficulties by making it difficult or impossible to override system actions once initiated. Highly automated systems are also increasingly complex (in terms of both number and function of components) as well as coupled to other systems and components. Complexity and coupling make these systems more difficult for the operator to understand and predict (Sarter, Woods, and Billings, 1997). These problems are compounded in systems which exhibit poor observability – that is, in systems that do not actively support the operator in monitoring and understanding machine actions and intentions (Woods, 1996). Highly automated systems create the potential for breakdowns in human-machine coordination when the human operator is unable to understand and predict machine intentions and actions beforehand or is unable to detect or intervene in undesired system behavior after-the-fact.

The Cali crash (Aeronautica Civil, 1996) is one recent example of a breakdown in human-machine coordination caused by an inability of the human operators to detect and resolve conflicting human and machine intentions and actions in a timely manner. In this accident, a Boeing B757 crashed into a mountain enroute to Cali, Colombia. During the descent into Cali, the crew was cleared by air traffic controllers to fly direct to a point called "ROZO" which was co-located with a radio beacon identified by the letter "R". When the crew instructed the automated systems to fly to ROZO, they entered the letter

"R" into the Flight Management Computer (FMC) (an automated system that, among other things, provides steering commands to the autopilot). Since the waypoint name "R" was shared by several other waypoints, the FMC recognized this ambiguity and requested the crew select the desired waypoint from a list of all waypoints named "R". The list was ordered so that the closest "R" was at the top. Unfortunately, the point "R" chosen by the crew (who were operating at night and under a high degree of time pressure) from the top of this list was not the desired point, but instead a point over 100 miles behind the aircraft's current path. Unbeknownst to the pilots, the desired waypoint was not in the database as "R", corresponding to the identifier of the co-located radio beacon, but instead was listed under the name "ROZO". The system design in the B757 dictates that once a change is entered into the FMC, the aircraft will not pursue the new target until the crew presses a separate "execute" button. Perhaps due to time pressure or an over reliance on automated systems, the crew did not detect the pending conflict between human and machine goals and actions. The crew pressed the execute button, and the aircraft turned towards the undesired waypoint and ultimately crashed into a mountain while the pilots were still attempting to determine why the aircraft was not proceeding towards the desired waypoint.

This incident underscores the need to better understand the processes and factors that influence an operator's ability to *detect and correct* discrepancies between human and machine goals and actions both during and after the consent process. In this case, system design attempted to incorporate a measure of error tolerance by requiring explicit pilot consent before the aircraft would act on changes to the FMC database. However, factors such as time pressure and an over-reliance on automated systems may have contributed to the crew's inability to detect the conflict between their goal and the machine's goal. While the crew did detect the unexpected and undesired behavior of the aircraft shortly after depressing the "execute" button, they were unable to intervene in time to avert the accident.

The method of automation management described above is called management-by-consent. It is one of the most prevalent automation management strategies in many current automated systems. In this approach, automated systems cannot take action until and unless the human operator grants his or her consent to that action (Billings, 1997; Wiener, 1985). While this approach allows the operator the opportunity to consider the desirability of machine goals and actions prior to their execution, the Cali accident illustrates that it will only be effective if the operator is able to provide *informed* consent (Billings, 1997). If the operator is unable to fully understand the nature, extent, and implications of proposed machine goals and actions, human-machine coordination failures in the form of automation surprises (Sarter, Woods, and Billings, 1997) are likely and may lead to catastrophic outcomes.

In order to avoid automation surprises and other coordination breakdowns, informed consent requires that the operator to both *detect* and *resolve* goal and task conflicts. While conflict resolution (negotiation) processes are vital to ensure coordination, this study will focus on the factors that affect the first stage of the process, an operator's ability to detect potential task and goal conflicts in the first place.

In the aviation domain, one example of a proposed management-by-consent system is data link, the digital air-ground transfer of information. Data link is intended to reduce radio frequency congestion and communication errors by digitally transferring air traffic control clearances and information to the cockpit and presenting the information on visual cockpit displays. In some proposed data link systems, following pilot approval, data link clearance information may be routed (gated) directly into onboard automated systems. Data link gating has been proposed in order to reduce operator workload and communication errors by eliminating the need for the operator to enter these instructions manually (Knox and Scanlon, 1990). While the ultimate goal of introducing data link is to improve the safety and efficiency of the national airspace system, the system may also impose new automation management and coordination tasks on the human operator which will provide new opportunities for error in the form of human-machine coordination failures (Ritchie, 1990).

The success of data link will depend, in part, on the ability of the system to support the pilot in providing informed (as opposed to perfunctory) consent prior to gating and executing ATC clearance information. The vast majority of ATC clearances will not result in goal and task conflicts; however, data link systems must reliably support the detection of the few conflicts that do occur.

Numerous studies of the general performance effects and the acceptability of data link systems have been conducted in recent years (for an overview see Kerns, 1994; Rehmann, 1997). Still, a number of research questions remain unanswered. For example, very few studies have examined the effects of gating, and most of those have primarily focused on transaction times and subjective ratings of workload resulting from the presentation of *acceptable* clearances. They did not study the performance and acceptability effects arising from the presentation of conflicting and problematic clearances. The two data link studies that have studied the effect of gating on the detection of conflicts with data link clearances have produced contradictory results. Hahn and Hansman (1992) found that gating improved the pilot's ability to detect conflicting clearances, while Logsdon (1996) found that gating resulted in poorer conflict detection. Also, in a recent review of data link research, Rehmann (1997) identified the need to investigate the type of information that should be gated directly to the Flight Management Computer, and the effects that this gating will have on crew awareness and coordination. To contribute to a better understanding of the above findings and unanswered questions, the purpose of this study is to examine how data link gating will affect the pilot's ability to detect problematic clearances and implementations and how data link displays and procedures can be designed to better support informed consent (i.e. consent based on pilot awareness of the implications of gating information to these systems).

In order to address these issues, we will first examine existing theories and research on human-machine coordination and supervisory control. Next, we will discuss possible effects of factors such as the nature of a conflict, time pressure, and trust on an operator's ability to detect conflicts with machine intentions and actions. Finally, we will discuss the design and functional properties of the automated cockpit systems that may contribute to the potential cost and benefits associated with implementing future data link systems. Also, the results of earlier data link research will be reviewed. Consideration of these various factors and findings will serve to design a study that will produce contributions to both the applied and theoretical literature on human-machine coordination and collaboration.

2.0 Human-Machine Coordination

Both human-human and human-machine teams must coordinate their goals and activities to avoid conflicts and inefficiencies. Coordination within successful teams is apparently effortless and may be invisible to outside observers (Malone and Crowston, 1990). Coordination failures, however, are often quite visible and can result from a number of factors. For example, studies of expert systems (e.g. Suchman, 1987; Roth, Bennett, and Woods, 1987) have described coordination problems caused by limited machine communication capabilities and machine reliance on brittle procedural models of the human operator. In complex high risk domains such as aviation, automation surprises such as the previously described Cali crash provide salient examples of the coordination difficulties created by factors such as time pressure and over-reliance on highly complex and autonomous machine agents.

In order to better support human-machine coordination in future systems, we must first understand the processes involved in coordination as well as the human and machine roles in highly automated systems. The following sections will describe and define these processes, the structure of the human-machine relationships in supervisory control systems, and coordination in management-by-consent systems. Finally, the implications of the inherent tradeoff between the workload reduction benefits of automation and the costs imposed by reduced pilot involvement will be considered.

2.1 *What is Coordination?*

As Jordan (1963) first noted over 35 years ago, humans and machines are not independent but, instead are complementary. They must work together to achieve desired system performance. Even the most highly automated systems still require the presence of a human operator to monitor system performance and intervene in the case of system abnormalities and emergencies (Sanderson, 1989). In order for humans and machines to work together to achieve system goals, they need to develop or engage in processes and activities that ensure coordination and avoid conflict. But what does coordination mean?

Coordination can be viewed either as a *cooperative process* which requires agents to flexibly and adaptively work together towards goal attainment or it may be viewed as a *conflict resolution process* in which interactive participation is not required - instead the focus is on detecting and resolving conflicting goals and actions. The cooperative view is being taken in most research on human-computer interaction and computer supported cooperative work (CSCW). For example, Zachary and Robertson (1990) discuss the communication activities required to coordinate goals which can range from completely shared to completely independent. They define human-machine *cooperation* as a collective activity oriented towards a specific goal. From this viewpoint, regardless of the degree to which goals are shared, mutual communication of goals and actions is necessary to ensure cooperation. Silverman (1992) also examined the interaction between humans and machine critic systems (systems that provide advice and suggestions) and found that successful human-machine collaboration requires two-directional communication of intent as well as the capacity for flexibility and adaptability on the part of both human and machine agents (see also Clarke and Smyth, 1993).

Unfortunately, in many domains machine agents possess limited communication and inferential abilities that severely constrain true cooperation among human and machine agents (Norman, 1990). As a result, machine agents are unable to share the responsibility for coordinating intentions and actions because of limited machine abilities and the dynamic nature of some domains. A variety of research projects (e.g. Coury and Semmel, 1996, Jones, Mitchell, and Rubin, 1990; Cha and Funk, 1997) are exploring means of allowing machines to share in coordination activities and associated workload. While some of these efforts have been quite successful (Jones, Mitchell, and Rubin, 1990), efforts in the aviation domain (Cha and Funk, 1997) have not achieved complete success due, in part, to the dynamic nature of the domain. As a result, it appears that human operators in this complex high risk domain will continue to be responsible for the activities and shoulder the costs associated with coordinating human and machine actions.

Because of these limited machine capabilities, a less cooperative approach to human-machine coordination may be more appropriate. For example, Symon, Long, and Ellis (1996) prefer the term *coordination* as opposed to cooperation due to the emphasis that the term coordination places on goal conflicts and conflict resolution as opposed to the mutual cooperation and interactive communication implied by cooperation. They view coordination theory (Malone and Crowston, 1990) as the most comprehensive treatment of these issues. Coordination theory defines coordination as "managing dependencies between activities." This section will discuss those areas of coordination theory that apply to human-machine coordination with a focus on identifying the processes by which human operators may manage dependencies (such as goal and task conflicts) in future data link systems.

Coordination theory describes coordination as comprising actors (agents), activities, goals, and interdependencies between goals and actions. Human and machine agents can hold a wide variety of potentially conflicting goals. For example, on a typical flight, pilot goals may include navigating from airport A to airport B, following air traffic control directives, following company policy, etc. Automated systems also hold a wide variety of goals. Most of these goals such as heading, airspeed and altitude targets are provided by the pilots. Some goals, however, are provided by the aircraft designers. For example, autopilots (and even aircraft control software in the most advanced aircraft) are programmed to fly above a minimum airspeed and below a maximum airspeed at all times. In addition to goals provided by the operator or designer, some machine goals may be provided by other human or machine agents. For example, data link will allow for direct communication between cockpit automation and ground-based human and machine agents (Prevot, Palmer, and Crane, 1997). Detecting and resolving the conflicts created by this additional source of system input will create new memory and attention demands for the pilot.

Malone and Crowston (1993) have identified several types of dependencies (see table 1). Many of these do not directly apply to the proposed data link system (e.g. producer/consumer relationships, transfer, usability, and design for manufacturability). In contrast, shared resources (arranging for adequate supplies for a given task), task assignments (allocating the time required to complete a task), and simultaneity constraints (synchronizing two tasks may not or must occur at the same time), and task/subtask dependencies may play a role. In systems such as data link, however, once the operator provides consent to higher level machine goals, the automated systems control the assignment and scheduling of the tasks required to meet the goals agreed to by the operator. The pilot can control shared resources, task assignments, and simultaneity constraints only by determining the acceptability of system goals (goal selection) as well as the tasks that will be required to accomplish the overall goal (task decomposition) prior to the consent decision. Thus, this study will focus on the area of task and subtask dependencies (conflicts) since goal selection and task decomposition are the primary means of controlling these other dependencies.

Table 1. Dependencies and coordinating processes (adapted from Malone and Crowston, 1993)

Dependency	Examples of coordination processes for managing dependency
Shared resources	"First come/first serve", priority order, managerial decision, market-like bidding
Task assignments	(same as for "Shared resources")
Producer / consumer relationships	Prerequisite constraints Notification, sequencing, tracking
Transfer	Inventory management (e.g., "Just In Time", "Economic Order Quantity")
Usability	Standardization, ask users, participatory design
Design for manufacturability	Concurrent engineering
Simultaneity constraints	Scheduling, synchronization
Task/ subtask	Goal Selection, task decomposition

2.2 *Automation Management and Supervisory Control*

Since the human operator will, in all likelihood, bear the responsibility for detecting and resolving conflicts between human and machine actions in data link systems, human roles and capabilities will constrain the opportunities and influence the effort required to detect and intervene with undesired machine intentions and actions. The observed coordination costs and benefits will depend, to a large extent, on the specific form of automation management strategy implemented in the design of data link systems. For example, requiring operator consent prior to each machine action may increase awareness of machine actions. However, at the same time, these potentially frequent interactions may impose additional attentional demands and workload costs.

In addition to the example above, research in human organizations as well as distributed machine systems also points out the impact of management strategies on coordination. For example, human organizations that are centered around specific functions often facilitate coordination within a function but increase coordination difficulties between functions (see Hughes, Ginnett, and Curphy, 1996 for a brief overview). In distributed artificial intelligence systems, hierarchical networks often increase the number of communications required but decrease the amount of shared knowledge required by individual nodes or agents. (cf. Fox, 1988; Malone, 1988; Steeb, Cammarata, Hayes-Roth, Thorndyke, and Wesson, 1988). In contrast to these studies in other domains, relatively little is known about the effects of various automation management strategies on human-machine coordination. Given the large asymmetry in abilities and coordination capabilities of human and machine agents (Norman, 1990), it is likely that coordination effects in human-machine systems may diverge from those observed in purely human and machine domains. The following section will discuss the potential implications of implementing a management-by-consent system on the generic human roles in supervisory control systems.

Humans and machines can share functions in many ways. Billings (1997) has developed a comprehensive list of different methods of implementing automation in the aviation domain, drawing on the work of both Sheridan (Sheridan and Verplank, 1978; Sheridan, 1997) and Wiener (1985) (see table 2). These automation management strategies cover the spectrum from fully manual to completely automatic control. In general, the authority and autonomy of automated systems increases as strategies approach full automatic control. Although examples of each of these automation management strategies can be found in complex high risk domains such as aviation, the two highest levels of shared control - management-by-consent and management-by-exception - are expected to be prevalent in new highly autonomous automated systems (Billings, 1997).

Table 2. A continuum of automation management strategies (Billings, 1997)

Automation Management Mode	Human and Machine roles
Autonomous Operation	Operation in accordance with instructions provided by system designers; no human attention or management required (human intervention may be impossible).
Management-by-exception	Automation possesses the capability to perform all required actions and will perform all actions unless the human operator takes exception by manually intervening or reprogramming automated systems.
Management-by-consent	Automation, once provided general goals, operates autonomously, but will not act until and unless human operator provides consent
Management by Delegation	Once human operator provides specific instructions, automated systems will follow those instructions unless it is not capable of executing them.
Shared Control	Human provides control inputs that are modified and shaped by automated systems.
Assisted Manual Control	Human operator provides control inputs that are implemented by automated systems.
Direct Manual Control	Human operator physically controls the system.

In accordance with human-centered design principles which call for pilot involvement with and authority over automated actions (Billings, 1997), present and proposed data link systems allow the pilot to exercise system control via an accept/reject decision prior to automatically gating data link information into the FMC CDU and MCP. Thus, data link system can be classified as a management-by-consent system in which automated systems are not allowed to act until and unless granted specific permission by the human operator. Once goals are selected or agreed to by the human operator, the machine carries out those goals using its own strategies. Compared to strategies such as management-by-exception, the management-by-consent approach places a relatively large degree of control in the hands of the human operator by granting him/her control over the execution of machine intentions and actions. However, as pointed out by Billings (1997), consent must be informed consent. If the operator does not fully understand automated systems goals, the methods that will be used to achieve these goals, or the implications of these goals and actions for future system performance, he/she is not effectively in control and it is likely that breakdowns in human-machine coordination will occur. For example, in the Cali crash (Aeronautica Civil, 1996), the pilots gave consent for the automated systems to pursue an ill-understood target (fly towards an unintended waypoint). In this case the crew exerted control without fully understanding the implications of requested machine actions, resulting in unintended and disastrous system actions.

A closer examination of human responsibilities and tasks in supervisory control systems reveals potential problems for the operator's ability to provide informed consent. Sheridan (1997) identifies five basic human roles in supervisory control: planning, teaching, monitoring, intervening, and learning.

In the planning role, the operator decides which variables to manipulate, develops criteria to assess system actions, and determines constraints on activities. The planning process provides the basis for instructing automated systems and monitoring subsequent system behavior. In the current (non data link) cockpit, upon receipt of an ATC clearance, the crew plans by determining which autopilot mode and FMC CDU or MCP input will be required to execute that clearance.

Once a plan is developed, the pilot "teaches" the automated systems by providing the appropriate targets/instructions to automated systems. After providing input to the automated systems, the pilot then monitors system performance to ensure the system is performing as expected. Monitoring refers to all activities involved in adjusting system performance in response to small deviations (trimming), as well as fault detection and diagnosis. In the current cockpit, the pilot relies primarily on information presented on the Primary Flight Display (PFD) and Navigation Display (ND) to monitor system performance. These instruments give indications of aircraft attitude, altitude, airspeed, and heading, as well as autopilot and autothrottle modes and command targets. The pilot determines whether/when it is necessary to intervene with machine performance (due to, for example, task completion, machine requests for assistance, or undesired system performance). Finally, based on the given plan, inputs to the system, system behavior, and interventions (if any), the pilot learns lessons that may be applied to system control in future situations.

The implementation of a management-by-consent approach may affect each of these five human roles. Currently, at the planning level, the pilot often has some choice regarding the desired method of implementing an ATC clearance. For example, the pilot may implement an altitude change providing input only to the FMC CDU, only to the MCP, or to both the FMC CDU and MCP. The introduction of management-by-consent data link systems may result in reduced flexibility. In all likelihood, data link gating of an

ATC clearance will be limited to only one implementation method chosen by the system designer. As a result, when a pilot decides whether or not to accept a data link clearance, he/she will have to not only consider the appropriateness of the content of the clearance, but also know about/anticipate now the system will execute and implement it.

The introduction of a management-by-consent approach will reduce operator workload and involvement associated with teaching automated systems. Instead of providing multiple inputs to the MCP and FMC CDU, the pilot will only need to provide one input - consent to higher level machine goals (e.g. "proceed direct to point XYZ"). This will potentially reduce both the workload and errors associated with manual data entry. However, it may further remove the pilot from the control loop and perhaps lead to "out of the loop" problems associated with increased levels of automation such as delayed and less accurate event detection (c.f. Wickens, 1992). It will create new workload and attentional demands by requiring the operator to make decisions regarding the suitability of potentially frequent proposed machine goals and actions prior to providing consent.

These changes have implications for system monitoring. Since monitoring performance is affected by overall workload (Sheridan, 1997), reducing the workload associated with planning and manual data entry may free up resources for the monitoring task. For example, a data link study by Hahn and Hansman (1992) attributed the observed improved detection of erroneous clearances to this workload reduction. On the other hand, monitoring is driven, at least in part, by pilot expectations and mental models (Moray, 1986; Sarter, 1995; Sarter and Woods, 1997). The reduced human involvement in planning and teaching data link systems may prevent the pilot from forming adequate expectations and thus leading to less effective monitoring.

Even if the human does detect undesired system performance, the timing of pilot intervention with the observed machine behavior may be inappropriate. For example, Sarter and Woods (submitted) have found that, in the case of management-by-exception systems, pilot detection of and intervention in cases of uncommanded and undesired automation behavior were often significantly delayed, and sometimes missed entirely. A similar effect may be observed in management-by-consent systems. Once consent is given to a proposed but unsafe or undesirable goal or action, the pilot may fail to monitor system behavior closely due to strong expectations of adequate system activities. This assumption is supported by the findings of an analysis of aviation incidents from the Aviation Safety and Reporting System (ASRS) by Mosier, Skitka, and Korte (1994). This study indicates that, once pilots have delegated tasks to automated systems (in some cases by consenting to machine actions), subsequent human monitoring is often insufficient to detect deviations from desired performance. Also, in order to intervene effectively in undesired performance, the pilot must not only detect undesired performance, he or she must also formulate a plan to correct undesired system performance. If an operator does not understand the intentions and actions embodied in a machine request for consent, this lack of understanding may delay development of a plan to intervene to correct undesired system behavior. For example, in the Cali accident, even though the pilots were aware that the aircraft was not proceeding in the appropriate direction, they were unsure of the cause of the problem, and thus delayed initiating corrective action.

In summary, implementation of a management-by-consent approach has the potential to affect substantially the various human roles in supervisory control systems. These changes may affect the ability of the human operator to coordinate human and machine goals and actions. In particular, data link gating may decrease the workload and errors associated with planning and instructing automated systems, which may, in turn, lead to improved monitoring and intervention. However, decreased planning flexibility and less pilot involvement may, in turn, lead to monitoring and intervention difficulties that may offset these advantages. Additionally, the management-by-consent implementation of data link gating imposes the requirement for the operator to make potentially frequent and effortful decisions regarding the acceptability of proposed machine goals and actions, sometimes during high-workload, high-risk phases of flight. The ability of the human operator to manage/supervise machine activities by detecting conflicts and intervening in machine performance will depend on a trade off between the costs and benefits associated with implementing data link gating.

2.3 *The Cost-Benefit Tradeoff Between Involvement and Workload Reduction*

Manual data entry workload, operator involvement, and coordination costs are important factors in the cost benefit tradeoff described above. There is ample evidence for the importance of these factors in the literature on human-automation coordination. For example, manual (as opposed to automatic) control of a system has been shown to result in superior monitoring performance only in situations where the information gain associated with active participation outweighs the associated workload costs (Liu, Fuld, and Wickens, 1993). Idaszak and Hulin (1989) found that active participation in a process control task resulted in better failure detection and diagnosis. While subjects reported a greater workload associated with active participation, this workload cost was offset by improved system knowledge and increased information processing activities attributed to active participation. Similarly, in manual control studies, Wickens and Kessel (1979, 1980) found that the feedback benefits associated with manual system control outweighed the workload costs and allowed for superior detection of changes in control order.

The benefits of increased involvement inherent in active control do not always outweigh the associated workload costs. For example, Hahn and Hansman (1992) found that pilots were better able to detect unacceptable ATC clearances when using data link systems that automatically gated data link information to cockpit systems. In this study the workload reduction benefits associated with gating allowed pilots more time and effort to concentrate on the evaluation of clearance acceptability. Additionally, subjective reports indicated that manual data entry did not increase pilot understanding of the implications of an ATC clearance. Also, Hilburn, Jorna, and Parasuraman (1995) found that the use of automated air traffic control aids increased an air traffic controller's ability to detect pilot failures to obey ATC clearances. This advantage was attributed to the reduced workload costs associated with the use of automated aids which freed controllers to concentrate on the monitoring process.

These results indicate that the nature of the task may have a large impact on the costs and benefits associated with automated systems such as data link gating. When active participation results in improved feedback or system knowledge (e.g. Wickens and Kessel, 1979; Idaszak and Hulin, 1989) it will tend to facilitate monitoring and event detection. However, when active involvement does not contribute to system knowledge (Hahn and Hansman, 1992) or the workload reduction associated with automated systems is large (Hilburn, Jorna, and Parasuraman, 1995), then introduction of automated systems may result in superior monitoring and event detection.

In addition to workload reduction and system feedback, other task and environmental factors may influence the cost benefit tradeoff as well. For example, Milewski and Lewis (1997) found that machine communication skills and the ability to control agent performance (before, during, and after the task) are important factors affecting coordination with software agents. And for human-human teams, Saavedra, Early, and Van Dyne (1993) showed that increasing the number and complexity of task and goal interdependencies between human group members increases the amount of coordination, communication, and cooperation required between group members.

In summary, a number of studies in different domains have identified factors that affect human-machine coordination. These include: operator workload and involvement, the nature of the task, machine communication abilities, and the ability of the operator to control and intervene with the behavior of automated systems. The following section will discuss the findings from one of the first studies to examine the implications of some of these factors, as well as features of current system design, for human-machine coordination in future data link systems.

3.0 Operator Preference for Management-by-Consent: Findings from an Automation Survey

In order to better understand the influence of a variety of task and system factors, Olson and Sarter (1998) conducted a survey of pilot preferences for and performance under various automation management strategies. They asked 206 glass cockpit pilots from two major US airlines to rank order and explain their preferences for five different data link system designs across fifteen flight scenarios. The five different system designs represented one completely automatic system and two implementations each of both management-by-consent and management-by-exception. The two implementations of these automation management strategies varied in terms of pilot control and were included to study the effects of pilot involvement and workload. The fifteen scenarios reflected variations of the above factors. In order to help explain pilots' preferences and learn about shortcomings in the design of current flight deck technology, pilots were also asked to describe their operational experiences and problems with existing automated systems which represent either management-by-consent and management-by-exception.

Not surprisingly, pilots expressed a strong dislike for the fully automatic option and an overall preference for a management-by-consent approach due to perceived greater pilot control over machine actions. High time pressure and workload, as well as low task criticality, caused a significant number (but not a majority) of pilots to support a management-by-exception approach in which pilots could only intervene with machine actions after the fact by reprogramming or manually overriding automated systems.

The same factors - time pressure and workload - also affected pilot preferences between the two implementations of the management-by-consent approach. One implementation allowed pilots to separately accept or reject various elements of the ATC clearance (route, altitude, airspeed, etc.). This implementation allowed for greater pilot control over system behavior, but at the cost of higher workload. In contrast, the other implementation only allowed pilots to accept or reject the clearance in its entirety, which gave pilots less control, but also imposed less workload. Pilots expressed a significant preference for the lower control (and also less workload) implementation of management-by-consent in scenarios characterized by high time pressure and workload, as well as low task criticality.

In addition to their preferences for various implementations of future data link systems, pilots were asked to describe problems with existing flight deck automation. In particular, the survey asked pilots two questions: Have you ever experienced a situation where the automation did less than or more than you expected? and b) Have you ever experienced the automation to be too difficult or too easy to override? These two questions served to gather information that may explain pilots' preferences and to identify benefits and disadvantages of specific existing implementations of either management approach.

A majority of pilots (78.2%) reported that automation had violated their expectations at some point. A further breakdown reveals an almost even split between experiences where the automation did more than expected (39 cases), less than expected (55 cases), or both (57 cases). In response to the second question regarding difficulties with overriding automated systems, 43.9% of pilots reported having experienced problems. The vast majority of those problems were caused by the automation being too difficult to override. An analysis of the examples provided by pilots in response to this question reveals that overriding automation is not a unitary phenomenon. In fact, difficulties reprogramming (as opposed to manually overriding) automated systems accounted for 75% of these examples. These results highlight the difficulties that may occur if a system does not support pilots in giving informed consent in the first place. If a pilot fails to detect a potential conflict before consent is granted, reprogramming automated systems after the fact may impose substantial workload costs and take too long to avoid negative consequences.

A further analysis of the examples provided in response to these two questions revealed that human-machine conflicts fell into several distinct categories: a) automation pursued an undesired goal (14.4%), b) automation achieved the desired goal but also performed an unexpected task (unexpected coupling) (5%), c) automation achieved only part of the desired goal (lack of coupling) (4.4%), and d) situations in which the automation had the appropriate goals, but did not prioritize them as desired by the pilots (4.4%).

In summary, this automation survey showed a general pilot preference for a management-by-consent approach. Since operator preference data do not always directly translate to performance differences (Andre and Wickens, 1995), this study will examine the impact of time pressure and the nature of the conflict on the operator's detection performance and thus his/her ability to grant informed consent.

4.0 Factors Affecting Informed Consent

As discussed earlier, research on monitoring and fault detection in automated systems as well as the automation survey described above indicates that the ability to grant informed consent may be influenced by several factors. The following sections will discuss further the three factors that will be included in the design of this study: type of human-machine conflict, time pressure, and operator trust.

4.1 Type of Human-Machine Conflict

Providing informed consent requires human operators to consider the presence and nature of dependencies or conflicts embodied in proposed machine actions. This section will discuss different possible forms of goal and task conflicts as well as discuss the likelihood that these conflicts will be detected.

Goals can be related in either a negative or a positive manner (Wilensky, 1983). When goals are negatively related, opposing goals are held between two agents resulting in goal competition and requiring some type of intervention to ensure coordination. When goals are positively related, two agents possess the same or overlapping goals. In the case of positive goal relations, coordination may be still required in order to coordinate subordinate tasks and subgoals.

The distinction between positively and negatively related goals can be further refined. Pilots responding to the previously described automation survey (Olson and Sarter, 1998) indicated that problems encountered with existing automated systems could be grouped into four general categories: 1) automation pursued an undesired goal, 2) automation achieved the desired goal but also performed an unexpected task (unexpected coupling), 3) automation only achieved part of the desired goal (lack of coupling), and 4) automation pursued the appropriate goals, but did not prioritize them correctly.

The situation in which automation pursues an undesired goal is an example of a negative goal relation. The remaining three situations are best categorized as situations involving positive goal relationships: the automation pursued concordant higher level goals, but implemented those goals in a manner inconsistent with pilot expectations or desires. For example, in some aircraft, changing the landing runway selected in the FMC (the desired goal) will also delete the desired vertical profile (unintended machine actions) (see category 2 in the previous paragraph).

In many cases, automated systems fail to correctly implement agreed-upon goals because they are not able to sense other operator goals, or they employ brittle procedural models of the human operator (Suchman, 1987). These limited machine abilities can force automated systems to make (sometimes incorrect) assumptions regarding the goals and plans of the human operator. For example, in the above case, the automated system assumes that pilots always need to create a new vertical profile when the landing runway is changed. Therefore, it automatically deletes the existing vertical profile when, in fact, the pilot wishes to retain it. In other instances, automated systems may recognize that they lack knowledge of required human goals and actions and, instead of making potentially invalid assumptions, they may simply stop and wait for further instructions. For example, if a newly assigned arrival procedure does not share a common waypoint with the existing route, the automated systems will not proceed to the first point on the arrival procedure unless given specific instructions. If not noticed by the pilot, these situations may result in automated systems failing to accomplish the overall operator goal. In summary, even though automated systems may be pursuing goals that do not conflict with the human operator, limited machine abilities to sense and update human plans and actions can lead to implementation strategies that conflict with human goals.

In data link systems, negative goal conflicts will arise when the instructions embodied in the clearance itself conflict with other pilot goals. Positive goal conflicts will arise when the clearance itself is acceptable, but automated systems choose and implementation method that will result in failing to achieve, or achieving more than the instructions contained in the clearance. In order to avoid confusion between these two conflict types, negative goal conflicts will hereafter be referred to simply as "goal conflicts" while the negative goal conflicts resulting from poor implementation will be referred to as "implementation conflicts."

Supporting informed consent requires consideration of the relative difficulty in detecting a given type of conflict. If one particular type of conflict can be expected to be more difficult to detect than other conflict types, display design or procedures could be tailored to provide operators with additional support in those cases.

Some insights into relative differences in conflict detection may be obtained from the literature on human error detection. There are well-documented differences in the relative detection rates for a variety of human error types. In general, errors of omission (errors resulting from the failure to take an action) are more difficult to detect than errors of commission (errors resulting from taking an incorrect action) (Reason, 1990). More specifically, mistakes (errors in intention formation) and lapses (losses of intention) are considered most difficult to detect, while slips (errors in intention execution) are detected quite easily (Reason, 1990). Error detection mechanisms may explain some of these detection differences.

Sellen (1994) studied the detection of everyday errors and identified the following error detection mechanisms - action based detection, outcome based detection, detection by a limiting function, and detection by others. Outcome based error detection - detection based on a comparison between results and expectations - is most closely related to a human operator's decision to consent to machine actions as well as the decision to intervene after consent has been given. In order for outcome based error detection to occur: 1) expectations must exist regarding the effects of plans and actions, 2) those effects must be perceptible, 3) the state of the world must be sufficiently monitored, and 4) the individual must associate any discrepancy between observed and expected effects with his or her actions (Sellen, 1994).

Studies of pilot-automation interaction (Wiener, 1989, Sarter and Woods, 1992, 1994, 1997, submitted) provide strong evidence of the effects of inadequate mental models and poorly formed expectations on the detection of and intervention in undesired system performance. Pilots in highly automated aircraft have switched from a regular, repetitive and automatic monitoring strategy (the basic instrument scan) to an expectation-driven approach in which pilots monitor system behavior primarily to verify that automated systems status and behavior match pilot expectations (Sarter and Woods, 1997). At the same time, we know that pilots often possess inadequate and incorrect mental models of the function and structure of automated flight deck systems. As a result, they find it difficult to predict or understand machine behavior and frequently experience automation surprises, especially when it does more than expected (Wiener, 1989; Sarter and Woods, 1992, 1994, 1997; Olson and Sarter, 1998). These effects are exacerbated by high time pressure or in abnormal situations (Sarter and Woods, 1994).

In the context of data link gating, both inadequate mental models and poor system observability may again contribute to delayed or failed conflict detection in several ways. First, while automated systems often clearly indicate their goals (by displaying system targets), less effective feedback is being provided on the method used to achieve these goals (e.g. modes and their implications). As a result of this poor observability, implementation conflicts may be more difficult to detect than goal conflicts. Second, gating may lead to delayed or failed conflict detection since it removes the pilot from actively planning and instructing automated systems. This decreased involvement in the planning process may lead to difficulties establishing mental models and expectations regarding machine behavior prior to a consent decision. Third, since monitoring is based on expectations, conflicts in which machines do more than expected may be more difficult to detect than those in which the machine does less than expected. During monitoring, pilot expectations guide allocation of attention to specific display indications in order to confirm that system performance matches expectations. In the absence of expectations, the detection of undesired system action depends on salient indications of undesired system performance. Given the low observability of many automated systems, detection of unexpected system response may be difficult. Since expectation-based monitoring is an effortful process, factors such as poor display design or time pressure that limit the operators ability to employ available mental resources to the monitoring process may contribute to detection failures.

4.2 Time Pressure

As evidenced by the Cali crash, time pressure can negatively affect human performance in a variety of ways (see Hockey, 1984 for a review). This section will discuss the effects of time pressure on decision making and monitoring and extrapolate those findings to anticipate the effects of time pressure on pilot detection of undesired clearances in data link systems.

Time pressure has been found to have significant effects on decision making in both naturalistic and laboratory settings (e.g. Wright, 1974; Rothstein, 1986; Wickens, Stokes, Barnett, and Hyman, 1993). In naturalistic decision making tasks, time pressure has been found to increase personal stress, shift cognitive process to less complicated reasoning strategies, and limit the options considered (Orasanu and Connolly, 1993). These general effects can manifest themselves in a variety of ways. Edland and Svenson's (1993) review of judgments and decision making under time pressure found that as time pressure increases: 1) subjects tend to use less information, or use available information in a more shallow manner, 2) more important attributes are given increasing weight, 3) subjects tend to lock in on one strategy, and, as a result, 4) accuracy decreases (although decrements may be delayed until time pressure becomes so severe that changes in information gathering and decision strategies no longer approximate more optimal solutions). It appears that changes in information acquisition precede changes in decision strategy as time pressure increases. Johnson, Payne and Bettman (1993) found that moderate time pressure resulted in decision makers reducing the scope of information considered while severe time pressure resulted in a qualitative change in the nature of the strategy employed. Under severe time pressure, operators were found to shift to less effortful heuristics such as elimination by aspects or the use of simple rules based on a small number of attributes.

In general, these studies suggest that as time pressure increases, operators will consider less of the available evidence in the consent decision and may also seek less cognitively demanding methods of arriving at the consent decision. These considerations, as well as work examining the cognitive effects of display design (e.g. Woods, 1995; Coury and Boulette, 1992; Rasmussen, 1986), suggest that displays which minimize the cognitive costs associated with extracting the relevant information may best support operator performance under time pressure. For example, displays that integrate information or increase the salience of the most relevant information may be helpful in supporting consent decisions under time pressure. These findings also suggest that time pressure should have the largest effects on those conflicts which are most difficult to detect. Conflicts that require the least amount of effort to locate and process relevant information may be least affected by time pressure.

While time pressure may decrease the likelihood that conflicts are detected during the initial consent decision, its effect on subsequent monitoring and intervention is not as clear. Liu, Fuld, and Wickens (1993) studied the effects of time pressure on monitoring a scheduling task. Time pressure was manipulated by asking subjects to report errors in response to a tone that occurred at twice the average reaction time or .75 times the average reaction time. The results showed that time pressure did not affect monitoring performance. However, other research suggests that time pressure may have some effect on error detection in supervisory control systems. Accumulation models of fault detection such Gai and Curry (1976) see error detection in supervisory control as a gradual accumulation of information over time. In this model, operators continue to sample system information over time until the cumulative likelihood that a fault is present exceeds a decision criterion. If the time interval is shortened without changing the decision criterion, error detection performance should decrease. It may be that time pressure will have a greater effect on more complex decisions that require the accumulation of greater amounts of evidence such as the decision to accept or reject a data link clearance.

In summary, time pressure has been shown to have large negative effects on decision-making processes similar to the consent process in data link systems. Increased time pressure will likely result in decreased cue sampling and simplified strategies that may, in turn, lead to decreased detection of conflicts with clearance goals or implementation. It is likely that factors such as conflict detection difficulty and display design will exacerbate general time pressure effects. In general, time pressure may affect monitoring processes following the consent decision to a lesser extent. However, as the complexity of the monitoring task increases, conflict detection at that stage may decrease under time pressure as well.

4.3 Operator Trust

Operator trust in automation has a significant impact on automation use and monitoring (e.g. Lee and Moray, 1992; Muir and Moray, 1996). The following section will define and describe trust as well as hypothesize about the effects it may have on the ability of the operator to detect erroneous clearances in management-by-consent systems.

Muir (1988, 1989) was one of the first to attempt to adapt theories of human trust to human behavior in automated systems. Muir (1994, p. 1911) defines trust as "…the expectation, held by a member of a system, of persistence of the natural and moral social orders, and of technically competent performance, and of fiduciary responsibility, from a member of the system, and is related to, but is not necessarily isomorphic with, objective measures of these properties." Muir also includes Rempel's (1985) three dimensions of trust development (predictability, dependability, and faith) in her complete model of trust. Muir and Moray (1996) empirically validated these six components of trust (persistence, competence, fiduciary responsibility, predictability, dependability, and faith) in a study of trust and operator behavior in a simulated process control environment. This study found that competence was most closely associated with the subjective ratings of overall trust in system components. These findings suggest that the perceived competence of the air traffic controllers who generate data link clearances, the data link systems that translate the clearance into a set of targets and commands, as well as the perceived competence of automated systems such as the FMS to achieve those targets may influence a pilot's decision to intervene in automated system behavior either during or after the consent decision.

How do trust and reliability affect human performance? Empirical studies have shown that operator trust can affect both the use and monitoring of automated systems. Studies employing a simulated pasteurization plant (Lee and Moray, 1992; Muir and Moray, 1996) manipulated the nature and magnitude of errors in system components. In these studies, subjects were allowed to either manually or automatically control a set of pumps to regulate the flow of product through the plant. The pumps exhibited errors that varied in type (constant vs. variable) and also in magnitude. The results of varying the magnitude and nature of pump error showed that even a small degree of variability reduced operator trust. However, as the magnitude of the error increased, trust became increasingly insensitive to additional increases in the magnitude of the error. These results indicate that trust can be reduced even by small temporary failures. Trust was found to correlate well ($r = 0.71$ overall) with use of automated features; trust was also negatively correlated with monitoring (as measured by checking pump output)($r = - 0.41$). These results imply that high operator trust in automated systems should lead to decreased monitoring and decreased intervention.

In addition to these process control studies, automation reliability has also been shown to affect human monitoring of automated behavior in other domains. Parasuraman, Molloy, and Singh (1993) asked subjects to perform both a tracking and fuel management task of which the fuel management task could be completed manually or automatically. When the reliability of the automated aid was variable (less trustworthy), subjects monitored the fuel management task more closely. Similar results were found by May, Molloy, and Parasuraman (1993) who found that monitoring in conditions of very low reliability (reliability approximately 25%) was superior to monitoring performance in the high reliability condition (reliability > 90%).

While these results suggest that trust and self-confidence have direct effects on performance, Riley (1989, 1994, 1996) proposes that operator behavior is a function of a number of other influencing factors such as workload, risk, and task complexity. Riley (1994) explored the relationship between trust, workload, uncertainty associated with task automation, and risk in computerized test bed of two simple tasks (letter-number classification and a one axis tracking task). Both college students and professional airline pilots served as subjects. Automation use was related to automation reliability. Airline pilots exhibited the same general patterns as student subjects, however they tended to use automation to a much greater extent than the student subjects, suggesting training or experience may also influence the effects of trust on automation use.

A synthesis of these results suggests that trust affects operator use and monitoring of automated systems. When automation is reliable, it will engender trust in automation that will lead to decreased monitoring of automated systems. These findings have several implications for detection of erroneous clearances in data link systems. It seems likely that high levels of operator trust will result in decreased monitoring following the initial consent decision. It is also possible that increased trust, especially when combined with other factors such as time pressure, may lead to reduced information gathering prior to the consent decision.

This study will examine the relationship between trust and error detection during both the initial consent decision and subsequent monitoring processes. It will also provide an opportunity to examine the influence of time pressure on the relationship between trust and operator behavior. Finally, this study will consider trust in both the air traffic controllers who generate the data link clearance as well as trust in the automated systems to carry out those clearances. Detection of conflicts at the goal level (clearance through weather, etc.) may be influenced by the result of unreliability associated with (and perhaps trust in) the air traffic controllers who issue the clearance, while conflicts arising from the way in which those goals are carried out (automated systems doing more or less than expected) may be influenced by the unreliability associated with (and perhaps trust in) the automated systems which implement the clearances.

5.0 Supporting Informed Consent via Display Design

Previous sections have described human-machine coordination as the management of dependencies between human and machine goals and actions. The ability of the human operator to detect these dependencies may be the result of a tradeoff between the manual data entry workload reduction and the possible reduced operator involvement. As discussed in the previous sections, this tradeoff may be affected by factors such as the nature of the potential conflict, time pressure, and trust. The design of datalink gating displays may also have a significant effect on the ability of the human operator to detect conflicting goals and tasks.

Displays serve the purpose of communication which is an important coordination process (Malone and Crowston, 1993). Compared to human communication abilities, machine communication abilities remain extremely limited. Machines are often unable to detect changing human goals and tasks (Suchman, 1987). They are also often unable to communicate what they do not know or adequately direct operator attention to information important to the current context (Norman, 1990). While some human-machine coordination problems arise from a lack of feedback, other problems are caused by information overload (Woods, 1996). In many human-machine coordination failures, the information required to determine automated system intentions and actions is present, but due to attentional limits is not correctly perceived by the operator. Data availability, the mere presence of data is not enough; instead, automated systems must support data observability. Data observability, i.e. support for the operator's ability to extract useful and relevant information (Woods, 1996), requires minimizing the cognitive effort required to extract meaning from the available information.

One method of achieving increased observability is to increase the intelligence of the human-machine interface. Several recent research efforts (e.g. Coury and Semmel, 1996; Cha and Funk, 1997; Rouse and Morris, 1987; Jones, and Mitchell, 1995) have been aimed at the development of an intelligent interface that will present and direct operator attention to relevant information in a context sensitive manner. While this approach is promising, much work remains to be done before intelligent interfaces can be implemented in the aviation domain.

Therefore light, this study will examine the effectiveness of two display designs that can be accommodated by existing technology – graphic and text displays. The following section will review the often-contradictory empirical findings comparing textual and graphic displays. The Proximity Compatibility Principle (Wickens and Carswell, 1995) will be employed as a framework to analyze the likely costs and benefits of graphic and textual displays in support of informed consent in data link systems.

5.1 Graphic vs. Text Displays

A number of studies have examined the relative speed and accuracy effects of graphic and text displays across a variety of domains and tasks. In some cases, subjects were asked to classify stimuli on the basis of two or more attributes (e.g. Coury and Boulette, 1992; Legge, Gu, and Luebker, 1989; MacGregor and Slovic, 1986), and make decisions based on combinations of these attributes (e.g., Spence and Parr, 1991; Schwartz and Howell, 1985) while others have asked subjects to diagnose and detect faults during a process control or troubleshooting task (e.g. Gillie and Berry, 1994; Coury and Pietas, 1989; Desauliniers, Gillan, and Rudisill, 1988). There is no consensus that one display type is superior to the other across conditions. Instead, it appears that observed effects depend on the specific type of display and on the nature of the task (Sorkin, Mabry, Weldon, and Elvers, 1991).

The studies cited above have examined many of the processes involved in the initial consent process and subsequent monitoring of data link systems. Overall, few *accuracy* differences were found between graphic and text displays for multi-attribute judgments (Spence and Parr, 1991), for event categorization (Coury and Boulette, 1992), for decisions made after combining a set of attributes (Schwartz and Howell, 1995), and for fault detection in a process control task (Coury and Pietras, 1989). In contrast to the lack of accuracy differences, Spence and Parr (1991) did find that graphic displays resulted in faster categorization.

Display effects were shown to be mediated by time pressure. Schwartz and Howell (1985) as well as Coury and Boulette (1992) found that graphic displays resulted in superior accuracy performance only when decisions and judgments were made under time pressure. Schwartz and Howell (1985) attributed at least some of these display effects to subjects' tendency to oversample information in text displays when under time pressure. Coury and Boulette (1992) found that, while graphic displays resulted in generally superior performance, uncertainty regarding systems state also affected performance. They found that a graphic (polygon) display resulted in most accurate performance under high time pressure and low uncertainty conditions. However, accuracy in the graphic display condition decreased with increasing uncertainty. The advantages of the graphic display were attributed to the presence of emergent features, and the capability of the graphic display to support integral processing of the displayed information.

It has been proposed that one major difference between graphic and text displays is that information presented in graphic displays is processed holistically, or in parallel (Goldsmith and Schvanveldt, 1984; Wickens and Scott, 1983). In accordance with the object file theory of attention (Kahneman and Treisman, 1984, Kahneman, Treisman, and Gibbs, 1992), graphic displays combine display features into one object which allows parallel processing of these features. Conversely, elements in text displays are thought to be processed serially, thus requiring more time to process (Coury, Boulette, and Smith, 1989; Pomerantz, 1986). When time or resources are limited, the serial processing inherent in text displays may produce performance decrements. However, in addition to processing of display elements, many other factors such as the nature of the task, the presence of emergent features, and other display features such as clutter and information access cost will influence resulting performance.

5.2 The Proximity Compatibility Principle and Emergent Features

As indicated by the often inconclusive comparisons between graphic and text displays, the assumed parallel processing advantage for graphic displays is only one of many factors that determine the speed and accuracy of performance with either display format. Ecological approaches that emphasize the use of emergent features (e.g. Bennett and Flach, 1992), as well as the proximity compatibility principle (Wickens and Carswell, 1995), describe the attentional and performance effects of other factors such as task and display properties. The basic problem addressed by these approaches is that display properties that tend to support the operator's ability to divide his or her attention between elements of information (integrative processing) may impose costs on the operator's ability to focus attention on one particular element and vice versa. In contrast, Bennett and Flach

(1992) suggest that, in many cases, graphic displays can support both focused and divided attention tasks without imposing a performance penalty. The following section will discuss the mechanisms that may underlie performance on these tasks and extend these theories to the use of graphic and text displays in a data link environment.

The PCP (Wickens and Carswell, 1995) is based to a large extent on Garner's work on dimensional integrality and separability (Garner, 1970, 1974; Garner and Felfoldy, 1970). The PCP posits that best performance will result from a match between display and task proximity. Tasks that exhibit close task proximity (require information integration) will benefit from displays that exhibit close display proximity while tasks which require little or no integrative processing (low task proximity) will benefit most from more separable (low display proximity) displays. Display proximity can be manipulated via six methods: spatial proximity, connections, source similarity, code homogeneity, object integration, and configuration (Wickens and Carswell, 1995). From the perspective of graphic vs. text displays, it is important to note that all display proximity manipulations, except object integration and configuration, can be employed by text displays.

The PCP states that the effects of combining a given level of display and task proximity will be mediated by four basic underlying information processing mechanisms: information access cost, object integrality, confusion and clutter, and emergent features (Wickens and Carswell, 1995). Information access cost refers to increased visual search time as a result of eye and head movements, which may be reduced by spatial proximity, feature similarity, connections, and enclosures. Since integrative tasks place a greater load on working memory, they will be affected to a greater extent by increased information access cost. Object integrality confers benefits via parallel processing as described by the object file theory (Kahneman and Triesman, 1984, Kahneman, Triesman, and Gibbs, 1992). Object integrality may provide benefits to both integrative and separable tasks unless response conflicts occur at a later processing stage. Confusion and clutter disrupt movement of attention and decrease discriminability of visual signals. Confusion and clutter impose costs on both high and low proximity tasks. Finally, emergent features are features of graphic displays other than those inherent in the raw codes (Wickens and Carswell, 1995). For example, triangular shape resulting from the combination of three variables is an example of an emergent feature. The presence of an emergent feature greatly simplifies the operator's task by turning an effortful judgment into simple pattern recognition (Bennett and Flach, 1992). Several studies (e.g. Sanderson, Flach, Buttig, and Casey, 1989; Sanderson, Haskell, and Flach, 1992) show that emergent features only support performance if they are meaningfully related to the semantics of the task (Bennett and Flach, 1992). Also, in the case of separable tasks, the salience of emergent features may decrease performance. However, Bennett and Flach (1992) contend that proper display design can eliminate these costs. Increasing the perceptual salience of elemental features may allow a display to support both focused and divided attention tasks.

The bottom line is that task performance may be the result of an interaction between task proximity and display proximity. Both graphic and text displays hold the potential to support both high and low proximity tasks by decreasing information access cost as well as clutter and confusion. Graphic displays alone, however, may employ object integrality and emergent features to support performance in integrative tasks (and potentially decrease performance in low proximity tasks). The advantages of graphic displays will lie in the employment of emergent features and object integrality while avoiding clutter and confusion.

In summary, the empirical evidence concerning graphic vs. text displays is mixed. For tasks similar to the detection of erroneous clearances, little benefit has been shown for graphic displays. It appears, however, that when combined with time pressure, advantages may be found due to the reduced effort required to extract information from graphic displays. Advantages for graphical displays may also accrue to the extent that these displays can invoke meaningful emergent features while retaining some degree of salience with regard to the display elements that may require separable processing. Text displays, on the other hand, may provide some benefits in integrative tasks by using properties such as color, spatial proximity, and grouping to reduce information access cost and decrease clutter and confusion.

6.0 Data Link Gating and Flight Deck Automation: The Domain Under Consideration

While the modern "glass cockpit" aircraft contains a number of automated systems, data link systems which transmit Air Traffic Control (ATC) clearance information will interact primarily with the two interfaces used to control the aircraft's vertical path, horizontal path, and speed - the Flight Management Computer Control Display Unit (FMC CDU) and the Mode Control Panel (MCP).

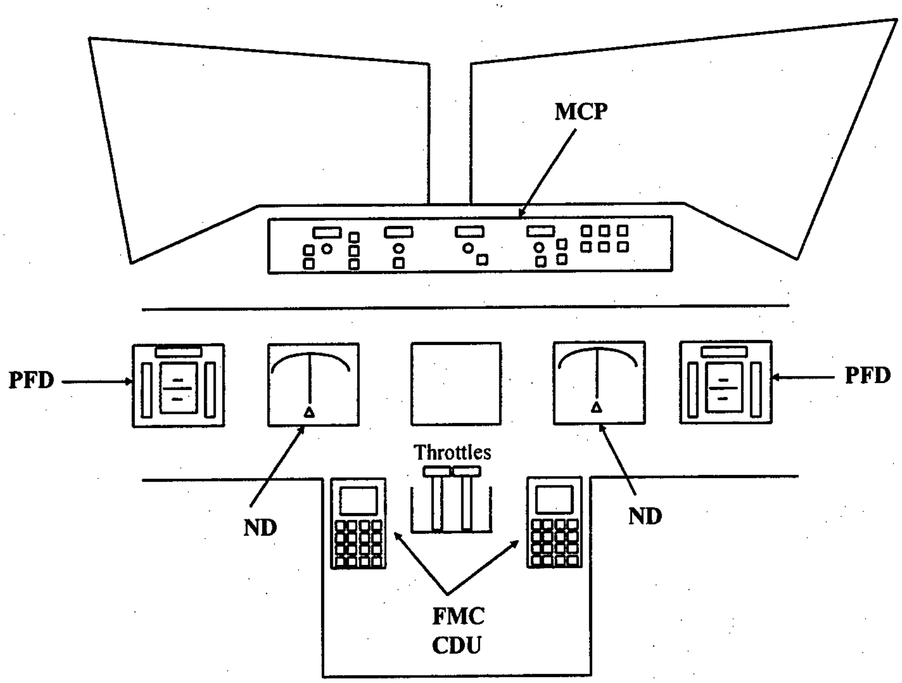


Figure 1. Typical layout of a glass cockpit aircraft.

In modern transport aircraft, hydraulic and electric actuators provide inner loop control functions by physically moving the control surfaces. These actuators, in turn can be controlled by the pilot either directly via the throttles and flight controls, or indirectly through the autopilot and auto throttle systems. During the vast majority of flight operations, pilots delegate basic aircraft control to the autopilot and auto throttle systems. The pilot controls these automated systems primarily through two interfaces, the FMC CDU and the MCP. Figure 1 depicts the typical location of these interfaces in the glass cockpit aircraft. The design and functional properties of the FMC and MCP will have a large effect on the coordination activities and problems pilots may encounter as a result of gating data link information to these systems. The following sections will describe the basic functions of these two interfaces, as well as their potential interaction with data link systems.

6.1 *The Flight Management Computer (FMC)*

The FMC Control Display Unit (FMC CDU) is the pilot's interface with a multifunction computer system (the FMC) that allows the pilot to plan, navigate, and control the aircraft. Through interconnections with a number of onboard systems and sensors, FMC planning features provide the pilot with weather (winds/temperature), fuel, timing, and performance data (optimal altitudes, takeoff and landing speeds, etc.). The FMC also contains a worldwide data base of navigational and instrument approach data that, when combined with satellite or inertial position information, allows the pilot to determine aircraft position as well as the relative position of other navigational waypoints. Finally, interfaces with the autopilot and automatic throttle systems allow the FMC (depending on mode) to provide steering, altitude and speed commands to these systems.

FMC CDU

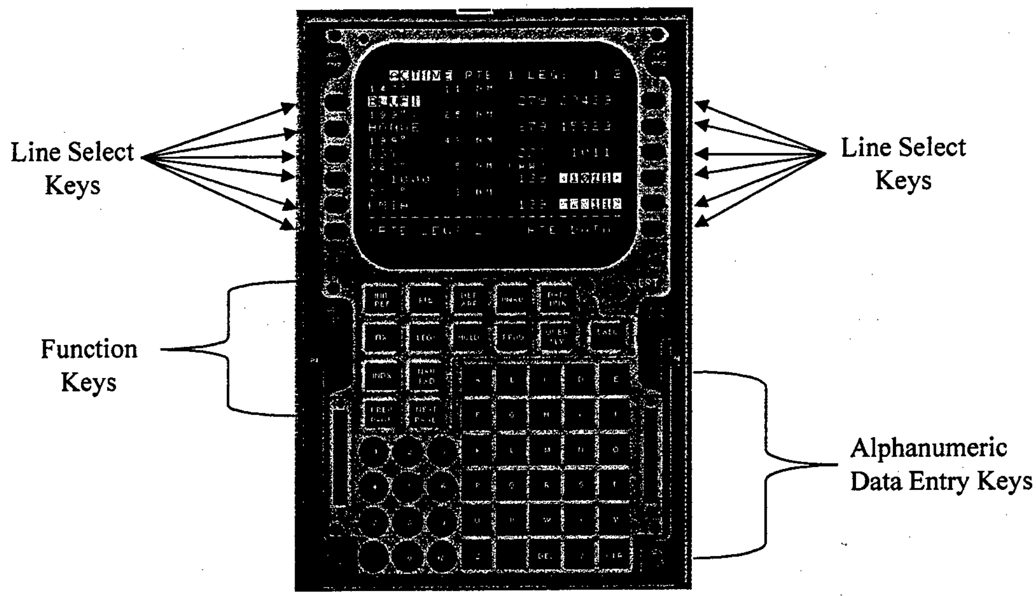


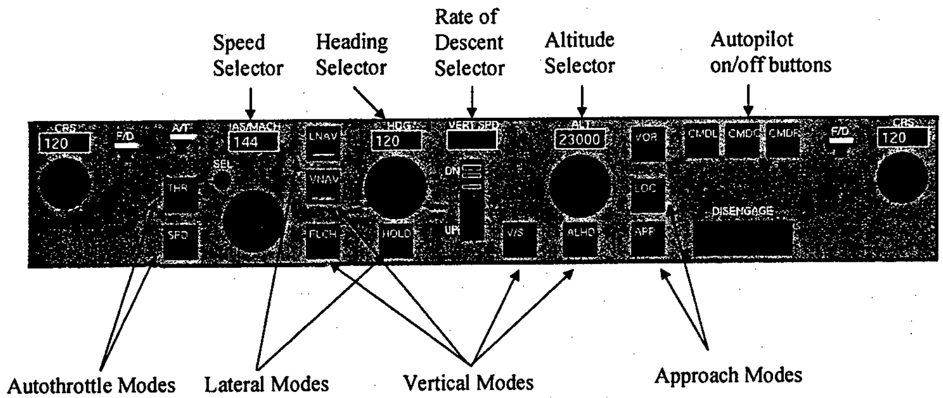
Figure 2. A typical FMC CDU.

The FMC CDU allows the pilot to input or review data via a menu driven architecture. Figure 2 represents the FMC CDU similar to the one the Boeing B757 aircraft. Data presentation is limited to approximately 12 lines of data arranged on either side of the display unit. In order to support the wide range of functions available, the FMC employs a branching menu structure in which pilots can access by selecting the appropriate function key (Legs, Route, Cruse, etc.) on the associated data entry panel. Once a given function is selected, the pilot can navigate through the associated menu pages by using the "prev page" and "next page" buttons. Although there are several different manufacturers, the underlying architecture, controls, and visual presentation are highly similar across different FMC CDU units.

The FMC CDU allows the pilot to input a desired route of flight, vertical profile, and speed profile. Route of flight information may be entered as waypoints (each flight is composed of a set of many waypoints) on the appropriate page of the FMC CDU via either manual keyboard entry or selection of pre-stored data base options via the line select keys adjacent to the display screen. Altitude constraints (either cruise altitude or a restriction to cross a horizontal waypoint or altitude at a given airspeed) may also be entered in the same manner. Aircraft speed may be controlled by either directly entering a speed value on the appropriate page, or by selecting a default speed profile (based on fuel economy or range considerations).

6.2 *The Mode Control Panel (MCP)*

The FMC CDU is not the only means by which the pilot can control aircraft speed, heading, and altitude. The MCP (see figure 3) allows the pilot to control autothrottle and autopilot modes, as well as to provide heading, altitude, air speed, and vertical speed targets to these systems. Autopilot and autothrottle modes are selected by depressing the appropriate buttons (e.g. LNAV, VNAV, FLCH, etc.), while airspeed, altitude, heading and vertical speed values are entered into the appropriate window via the associated selector knob. Although the distinction is not perfect, the FMC CDU is considered a "strategic" interface while the MCP is considered a "tactical" interface (Billings, 1997). The FMC CDU is often used to implement actions that will take place or continue relatively far into the future (e.g. entering changes to the route of flight), while the MCP is often used to implement more immediate actions such as flying an assigned heading or climbing to a given altitude. Like the FMC, there are differences among manufacturers and models. However, at a conceptual level most MCP functions are very similar.



Mode Control Panel (MCP)

Figure 3. A typical Mode Control Panel (MCP).

In order to control aircraft performance via the MCP, the desired target(s) must be entered into the appropriate window(s), and the appropriate mode(s) must be selected. For example, in order to comply with the clearance "fly heading 180°", the pilot must set 180 in the heading window and select the heading mode by depressing the top of the heading selector knob. There is a significant degree of coupling between the FMC CDU and MCP, as well as between autopilot modes. Some autopilot and autothrottle modes automatically activate other associated modes, while some information entered into the FMC CDU will not be acted upon unless the appropriate autopilot mode is selected on the MCP. For example, LNAV (lateral navigation) and VNAV (vertical navigation) modes must be selected on the MCP in order for the autopilot to follow the horizontal and vertical guidance commands entered into the FMC. Additionally, in some cases system behavior depends on the values set in both the FMC and MCP. For example, when descending in the VNAV autopilot mode, the controlling altitude will be the highest of either the altitude set in the MCP or an altitude restriction set in the FMC.

The complexity and coupling of modes and functions of the FMC and MCP contributes the attentional and knowledge demands these systems place on the human operator. A number of studies (e.g. Sarter and Woods, 1992, 1994, 1997, submitted) as well as incidents (e.g. Nagoya - Sekigawa and Mecham, 1996; Toulouse - Aviation Week & Space Technology, 1995) have highlighted the pervasiveness and severity of mode awareness problems. While data link holds the potential to improve air-ground communication, it also holds the potential to merely replace communication errors with other error types (Ritchie, 1990). In particular, the increased complexity and coupling of data link systems may make it more difficult for pilots to understand and predict the effects of gating a data link clearance, thereby contributing to poor detection of conflicts between human and machine goals and actions.

6.3 *Data Link Gating*

Data link is the proposed medium for two-way digital transfer of a variety of information between the cockpit and ground based systems. This study focuses on the transmission of Air Traffic Control (ATC) clearance information from ground based air traffic control systems to the cockpit. Data link clearance information can be uplinked and presented on either cockpit displays or printouts. In some proposed system implementations, the information may also be sent, or "gated", to the FMC CDU or MCP. For example, in the current non-data link environment, if a crew is given the clearance "proceed direct to point XYZ, climb and maintain 3,000 feet," this clearance would be transmitted via voice over a radio link and the crew would return a verbal acknowledgment and acceptance of the clearance. The crew would then manually enter "XYZ" into the FMC CDU as the next waypoint and enter the value 3,000 into the altitude window on the MCP. In a data link system, the clearance will be transmitted digitally and may appear on either a cockpit display or printout. The pilot may also electronically transmit acknowledgment and acceptance by pressing the appropriate button on the data link interface. In data link systems that do not incorporate gating, the pilot will still be required to manually enter the clearance information into the FMC CDU and MCP. In data link systems that incorporate gating, once the pilot accepts the clearance, the targets XYZ and 3,000 will be automatically entered into the FMC CDU and MCP respectively.

Data link gating has been proposed as a means of reducing the workload associated with the manual data entry as well as a means of minimizing data entry errors (Knox and Scanlon, 1990). While gating data link clearance information may fulfill the promise of reduced pilot workload and data entry errors, it may create other problems. Data link gating may add to the already existing problems associated with system autonomy, authority, and complexity, and coupling in glass cockpit aircraft. It may further reduce pilot involvement with, and thus knowledge of and ability to predict, the resulting machine actions and interactions with coupled systems. These effects may increase pilots' difficulties with detecting and resolving conflicts with machine goals and tasks. The following section will review a number of data link studies that have examined a variety of gating issues.

7.0 Earlier Research on Conflict Detection with Data Link Gating

Data link systems and procedures have been widely studied in the last decade. A recent review of the data link literature (Rehmann, 1997) cites over 250 studies, reports, and guidelines covering a variety of technical and human issues related to data link systems. Since the focus of this study is on the effects of data link gating on the pilot's ability to detect conflicts with machine goals and tasks, only a subset of these 250 reports are directly relevant. And of the over 40 studies and reports that have examined issues related to gating data link information (see Rehmann, 1997 for a comprehensive list), only a small subset have empirically examined the results of gating on pilot performance (Groce and Bocek, 1987; Knox and Scanlon, 1990; Waller, 1992; Chandra and Bussolari, 1991; Hahn and Hansman, 1992, Lozito, McGann, and Corker, 1993; Van Gent, 1995; Logsdon, Infield, Lozito, Mackintosh, McGann, and Possolo, 1995; Logsdon, 1996). The majority of these studies have focused on issues surrounding pilot-controller coordination and system acceptability by examining pilot responses to *acceptable* ATC clearances. The primary measures employed in these studies include response time, pilot workload, subjective acceptability, and situation awareness. In general, these studies have found that pilots consider data link gating to be a desirable feature (e.g. Van Gent, 1995, Hahn and Hansman, 1992, Waller, 1992). However, mixed findings were reported with respect to performance with datalink, especially in terms of response time (e.g. Waller, 1992; Van Gent, 1995). Also, these studies have examined the ability of different datalink displays to support pilot responses primarily to acceptable clearances (e.g. Van Gent, 1995; Lozito, McGann, and Corker, 1993; Waller, 1992).

Only very few studies (Chandra and Bussolari, 1991; Hahn and Hansman, 1992, Logsdon, 1996) have directly addressed human-machine coordination issues associated with data link gating in the context of unacceptable and erroneous clearances. Two of these studies (Chandra and Bussolari, 1991, Hahn and Hansman, 1992) have found that gating resulted in improved detection of erroneous clearances. Logsdon (1996), however, found that gating was associated with reduced detection of erroneous clearances. The following section examines possible reasons for these conflicting results.

7.1 *Gating Supports Conflict Detection*

Since Hahn and Hansman (1992) essentially replicated and expanded the original work in the same MIT lab by Chandra and Bussolari (1991), this section will describe the goals, methods, and manipulations in the Hahn and Hansman (1992) study only. This study was intended to examine the effects of data link gating, display method, and clearance readback on the pilot's ability to detect erroneous ATC clearances. Data link information was either gated automatically to both the FMC CDU and MCP, or entered manually by the pilot. Clearance information was either presented verbally via a simulated radio transmission, textually by presenting a verbatim copy of the verbal clearance on an additional visual data link display, or graphically by depicting routing changes on the map display and indicating changes in airspeed, altitude, and heading on the primary flight display. Additionally, in some conditions, pilots were required to read back, or repeat, the clearance while, in other conditions, they were not. Each subject saw all combinations of these three variables (except for the manual programming/no readback condition which was only presented verbally). This design resulted in each subject participating in 10 different experimental conditions.

Data was collected on nine B757/767 qualified airline pilots who performed the tasks of the pilot flying the aircraft (PF) over a set of 10 different scenarios (one for each of the 10 experimental conditions) using a desk top "glass cockpit" flight simulator. Each scenario simulated a descent and arrival (approximately 20 minutes in duration) through heavy weather to an airport in the Northeastern US. During the course of each scenario, the subject received 5 data link clearance amendments, two of which were erroneous. One erroneous clearance in each scenario involved a routing error (a clearance to an inappropriate waypoint, to an incorrect destination, or vector away from the desired path), while the other error represented a clearance into dangerous weather conditions. Dependent measures included detection of erroneous clearances, detection time, as well as subjective measures of effectiveness, time efficiency, and situation awareness.

Compared to both text and verbal formats, graphical displays resulted in more reliable and quicker detection of erroneous clearances. In spite of this performance advantage, the text display was subjectively rated as more efficient due to the consolidation of clearance information in one display (in the graphic condition, information was presented on the map display as well as the primary flight display). Gating was associated with superior error detection rates for routing errors. Subjects in the gating condition detected an average of 64% of erroneous routing clearances, while subjects in the manual loading condition detected only 42%. Detection of weather related errors was near perfect in all conditions resulting in no significant error detection effects due to gating. The gating benefits observed for detection of routing errors were attributed to the associated reduction in manual data entry workload that allowed pilots more time and resources to evaluate the implications of accepting the data link clearance. Pilot comments also indicated that the increased involvement associated with manually programming the FMC and MCP did not help them understand the implications of accepting a clearance. Note that these results of this study must be interpreted with caution. The relatively small sample size and lack of statistical analyses raise concerns over the robustness and generalizability of these findings.

7.2 *Gating Inhibits Conflict Detection*

Logsdon (1996) examined the effects of data link message length (long vs. short) and gating capability (automatic vs. manual FMC and MCP loading) on transaction time and erroneous clearance detection. Message length and gating capability were presented in a within subjects design, resulting in each subject participating in four different experimental conditions.

11 current glass cockpit pilots flew a set of four scenarios (one in each experimental condition) on a desktop simulation of an advanced transport aircraft. Subjects also performed a secondary compensatory tracking task in order to generate additional workload. The four scenarios consisted of two flights from San Francisco to Sacramento and two return legs. Transaction time, error detection, and subjective preference data were collected. Each leg contained four or five data link clearances, one of which was erroneous. Four types of erroneous clearances were employed. One was a clearance through weather, one was a vector in the opposite direction of final approach course, one was a clearance to descend to an altitude above current altitude, and one was a clearance to increase airspeed to a value below current aircraft speed.

Not surprisingly, when compared to manual loading conditions, gating resulted in significantly shorter total transaction time. Clearance length had no effect on either transaction time or error detection rate. In contrast with Hahn and Hansman (1992), however, gating resulted in a significant *decrease* in detection of erroneous clearances (65% detected in manual condition vs. 27% detected in the gating condition). In spite of these results, the majority of subjects (64%) felt that gating increased their situation awareness.

Several possible explanations may account for the different findings of these two studies. First, as demonstrated by the large difference between detection of weather and routing errors in the Hahn and Hansman (1992) study, the nature of the erroneous clearance may have a large affect on detection performance. This may be especially true since two of the erroneous clearances presented in Logsdon's study, the clearance to descend to an altitude above current altitude and the clearance to slow to a speed above current air speed, represent an error type (an "impossible" conflict) that was not included in Hahn and Hansman's study. "Inappropriate" conflicts such as a clearance through weather represent a conflict between the goals embodied in the clearance and other goals held by the pilot. In contrast, the impossible conflicts presented in Logsdon's study represent a conflict between the semantics of the clearance (i.e. "descend to") and the current state of the aircraft. The information required to detect an impossible conflict can be more difficult to present in a data link display and may require more effortful processing. This may explain the poor detection performance in Logsdon's study. This study will include impossible conflicts in addition to goal conflicts in which the instructions embodied in the clearance conflict with other pilot goals or flight procedures in order to examine this possibility.

Another difference between the two studies that may explain the observed performance discrepancies is that Logsdon (1996) employed a secondary tracking task while Hahn and Hansman did not. Finally, differences between the datalink display in the two studies may have contributed to the reported findings. Logsdon (1996) employed a text display only while Hahn and Hansman's study contained both graphical and text display conditions. A closer examination of Hahn and Hansman's (1992) data reveals that, although statistical significance is not mentioned, subjects using the graphic display detected almost 80% of erroneous routing clearances while subjects using the textual display detected only approximately 50% of these errors. Additionally, detection of routing clearances by subjects in the text display condition was almost always substantially delayed. Unfortunately, Hahn and Hansman (1992) do not compare performance effects of manual loading vs. gating within the text display condition. While these studies by Hahn and Hansman (1992) and Logsdon (1996) show that data link gating can have a large impact on the pilot's ability to detect erroneous data link clearances, they do not help us understand what factors influence the nature and magnitude of these effects.

8.0 Hypotheses and Expectations

As the preceding sections have shown, little is known about the impact of data link and gating on the pilot's ability to detect conflicts with goals embodied in a clearance or with the implementation of a clearance, both during the initial consent decision and during subsequent monitoring. Most data link studies (e.g. Knox and Scanlon, 1990; Waller, 1992; Van Gent, 1995) have examined only pilot responses to acceptable data link clearances. And the only two studies (Hahn and Hansman, 1992; Logsdon, 1996) that have examined conflict detection paint a contradictory picture of the effects of data link gating. Hahn and Hansman (1992) found that gating improved conflict detection while Logsdon (1996) showed the opposite effect. And finally, in addition to assessing performance outcome measures such as response speed and accuracy, this study will examine the effects of gating, time pressure, trust, and display design on the cognitive processes involved in conflict detection.

Our review of the literature suggests these apparently contradictory findings may be explained by the nature of a conflict, time pressure, operator trust, and display design. In the following sections, predictions will be made regarding the likely nature and direction of performance differences in response to these factors.

8.1 *Nature of the Conflict*

Based on the results of work on human error detection (Sellen, 1994; Reason, 1990) as well as studies of pilot interaction with automated cockpit systems (Wiener, 1989; Sarter and Woods, 1992, 1994, 1997, submitted), it is likely that some conflicts will be more difficult to detect than others. In general, goal conflicts should be easier to detect than conflicts arising from clearance implementation since the goals inherent in a (data link) clearance are more observable and may require little cognitive processing beyond understanding the content of the clearance. In contrast, conflicts due to clearance implementation require the operator to make assumptions about the translation of the clearance by the data link system into a set of commands that must be interpreted by yet another agent (the FMC). The additional effort and the reduced observability associated with identifying conflicts at the implementation level should contribute to less accurate and delayed detection of implementation conflicts compared to the detection of goal conflicts.

Differences in conflict detection speed and accuracy may also be observed between the two types of implementation conflicts presented in this study. Since pilots are known to monitor automated systems primarily based on their expectations of system activities (e.g., Sarter, 1995; Sarter and Woods, 1997), it is predicted that implementation conflicts in which automated systems do more than expected may be more difficult or take longer to detect than implementation conflicts in which automated systems do less than required by the goals embodied in the clearance. The decreased ability to detect implementation conflicts in which automated systems do more than expected should be evident in both the initial consent decision as well as in subsequent monitoring since, in both cases, information search and acquisition are driven primarily by pilot expectations.

We can make few strong predictions for the detection of impossible clearances (clearances in which instructions are impossible given the current aircraft state – e.g., a clearance to descend to an altitude above the current altitude). To the extent that impossible clearances require more effortful semantic processing, impossible clearances may be more difficult to detect during the consent decision than other goal conflicts, especially under high time pressure. Also, in some cases, the window of opportunity for detecting an impossible conflict after the consent decision is limited since the conflict will no longer exist once the aircraft attains the new desired performance targets.

8.2 *Time Pressure*

Time pressure is expected to decrease the pilot's ability to detect goal and implementation conflicts by forcing him/her to switch to less effortful and perhaps less accurate and comprehensive processing of the clearance. Under increased time pressure, pilots are expected to sample less information and perhaps switch to less effortful, simplified strategies (Edland and Svenson, 1993; Johnson, Payne and Bettman, 1993; Barnett and Wickens, 1986). As a result, during the consent decision, fewer goal and implementation conflicts should be detected. Both the verbal protocol data and the record of FMC CDU button presses should reflect a decrease in the information used to make the consent decision, and may indicate a switch towards simplified decision rules and strategies such as heuristics. Due to the added effort required to detect the less observable implementation conflicts, time pressure will likely have an even greater effect on the speed and accuracy of detecting those problems. Finally, time pressure may interact with display type. To the extent that the graphic display will allow pilots to rely on relatively effortless pattern recognition instead of more effortful integrative processing, time pressure should have a greater negative impact on the speed and accuracy of conflict detection in the text display condition.

8.3 *Operator Trust*

Operator trust has been shown to correlate highly with operators' use and monitoring of automated systems (Lee and Moray, 1992; Muir and Moray, 1996; Parasuraman, Molloy, and Singh, 1993; May, Molloy, and Parasuraman, 1993). In the case of datalink gating, we need to consider three different aspects of trust: trust in air traffic controllers to provide acceptable clearances, trust in data link systems to load the appropriate targets into the FMC CDU and MCP in a reliable and appropriate manner, and trust in automated systems to implement and pursue those targets in an adequate manner. Pilot trust in these three areas may be related to conflict detection both during the consent decision and during the subsequent monitoring. Previous research on trust and monitoring (e.g. Parasuraman, Molloy, and Singh, 1993; May, Molloy, and Parasuraman, 1993) indicates that conflict detection during monitoring should be negatively correlated with trust in both data link and automated cockpit systems. The effects of trust on conflict detection during the initial consent decision are less clear. High levels of trust may lead to a less thorough and/or less comprehensive review of clearance acceptability and implementation, and thus to the detection of fewer conflicts. Furthermore, detection of goal conflicts may be most highly correlated with trust in the air traffic controllers who generated the clearance, while detection of conflicts at the implementation level may be most closely associated with trust in the data link system's ability to load targets into the FMC CDU and MCP in an appropriate manner as well as trust in the ability of the FMC to execute those targets as intended.

8.4 *Display Design*

In the two gating conditions, the advantages of one display type - text - over another – graphic - will depend on the extent to which the display can minimize the cognitive effort required to extract relevant information. While the literature on relative benefits of graphic and text displays does not show a consistent advantage for graphic displays, the Proximity Compatibility Principle (Wickens and Carswell, 1995) can explain some of the earlier contradictory results. The PCP predicts that displays that provide the best match between task and display proximity should produce the best performance. The match between display and task proximity is mediated by four factors: information access cost, confusion and clutter, emergent features, and display integration. Although it may be difficult to classify the proximity of complex tasks (Bennett and Flach, 1992), general predictions can be made with respect to the four mediating factors described above.

In this study, it is expected that the lines depicting the current and the proposed vertical and horizontal profiles may produce emergent features that may allow the pilot to detect changes to the vertical and horizontal profiles via relatively effortless pattern recognition. However, this advantage may be offset to the extent that the highlighting of changes in our text display facilitates change detection. Also, given the relatively greater dispersion of information in the graphic display as well as the depiction of other navigation and weather information on the ND, the graphic display may increase information access cost as well as clutter and confusion. Finally, time pressure may affect performance differences between the two display conditions (e.g. Coury and Boulette, 1992; Schwartz and Howell, 1985). In general, under time pressure, displays that minimize the cognitive costs associated with extracting the relevant information should best support operator performance (e.g. Woods, 1995; Coury and Boulette, 1992; Rasmussen, 1986). Thus, the conflict detection speed and accuracy benefits of the graphic display should be greatest for urgent clearances.

9.0 Method

9.1 *Design*

The independent variables in this study were gating, display type, conflict type, and time pressure (see figure 4). Gating and display type were manipulated in a between-subjects manner resulting in three experimental groups.

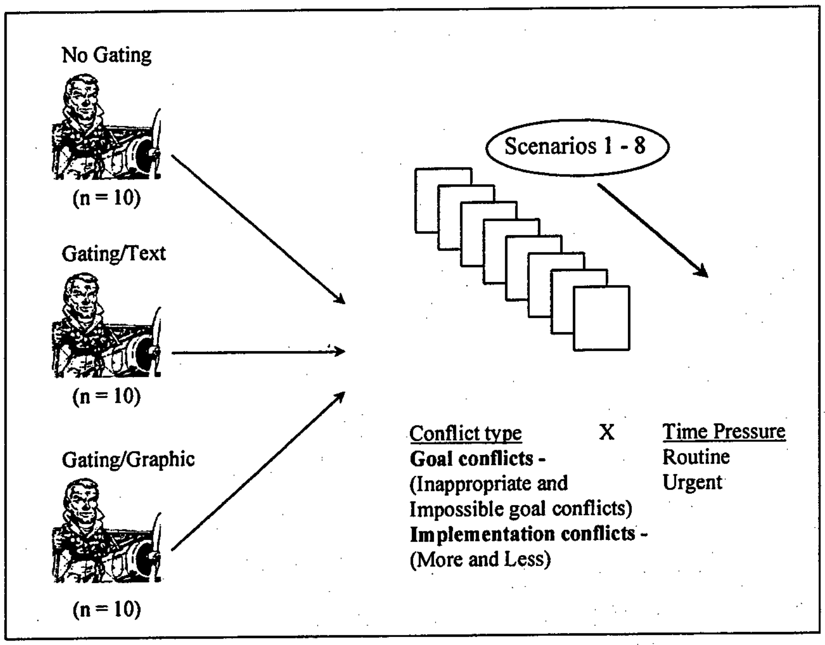


Figure 4. Experimental design.

The no gating condition served as a baseline and employed display only (no gating) data link procedures. In this condition, pilots were required to manually enter all clearance information into the MCP and FMC CDU. In the two gating conditions, after pilots read the text of the data link clearance, they could activate a "load" button to automatically transfer data link clearance information into the FMC CDU and MCP. In addition to the displayed text of the data link clearance, pilots in the gating/text display condition were presented with a text display highlighting the changes to performance targets that would result from loading a gated data link clearance. Pilots in the gating/graphic display condition viewed the same information in a graphic format. The specific details of the data link system and displays will be discussed in greater detail in section 9.4.

Conflict type and time pressure were varied in a within subjects manner across eight experimental scenarios. *Conflicts* embodied in data link clearances were one of two basic types: a) goal conflicts in which the goals embodied in the clearance conflicted with other pilot goals or the current system state, or b) implementation conflicts in which an acceptable clearance was loaded by the automatic data link system in a manner that, in combination with design features of the FMC, resulted in undesired aircraft behavior. Both goal and implementation conflicts were further broken down into two subtypes resulting in four different types of conflicts. These conflicts were: 1) inappropriate goal conflicts where the goals embodied in the content of the clearance conflicted with other pilot goals (e.g. an assigned speed restriction was too fast to allow for deployment of gear and flaps), 2) impossible goal conflicts where the instructions contained in the clearance could not be executed given the current state of the aircraft (e.g. a clearance to descend to an altitude above current aircraft altitude), 3) "implementation - does less" conflicts where the clearance itself was acceptable but was poorly implemented by automated systems, resulting in a failure to accomplish all of the goals specified or implied in the clearance, (e.g. a speed restriction given in the cruise phase of flight failed to propagate to the descent phase) and 4) "implementation - does more" conflicts where the clearance itself was acceptable but was poorly implemented by automated systems, resulting in inappropriate actions taken beyond the goals embodied in the clearance (e.g. a change to the horizontal route also deleted the vertical profile).

Time pressure was manipulated by presenting clearances at two levels of urgency – routine and urgent. For routine clearances, participants were asked to respond as soon as the situation and other tasks allowed. For urgent clearances, pilots were instructed to respond within 15 seconds in order to avoid a potential traffic conflict. Each type of conflict was presented at both levels of urgency during one of the eight experimental scenarios. Scenario order was counterbalanced within each group using a Latin square design.

9.2 *Participants*

Participants were 30 currently qualified B757 pilots from a participating major US airline. 10 pilots were assigned to each experimental condition in a quasi random manner. Pilots in these three conditions were matched on the basis of total flight experience, experience in the B757, and crew position (captain or first officer). Participation was voluntary, and pilots received 100 dollars for their participation in the study. Table 3 indicates the average experience level and crew position of pilots in each experimental condition.

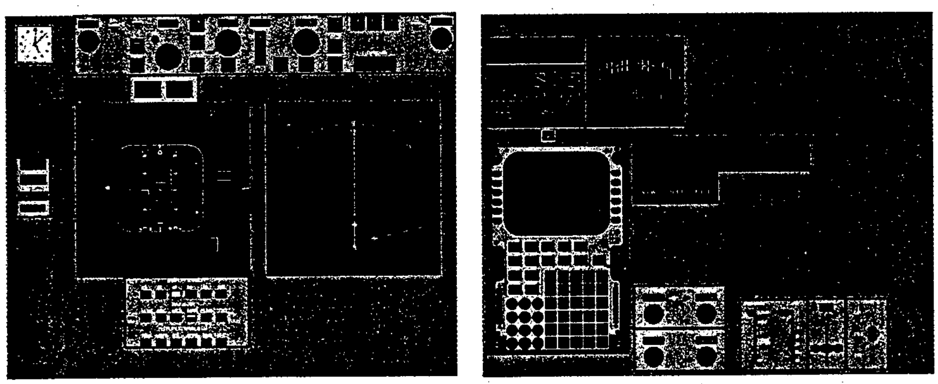
Table 3. Average pilot experience and crew position

Condition	Total time Avg hrs (range)	757 Time Avg hrs (range)	Captain/First Officer
No Gating	9410 (3,000 – 25,000)	1510 (300 – 4,300)	4 Capt/ 6 FO
Gating Text	10810 (3,100 – 21,000)	1743 (80 – 4,000)	4 Capt/ 6 FO
Gating/Graphic	10950 (4,000 – 25,000)	1922 (20 – 5,600)	5 Capt/ 5 FO
Average/Total	10390	1725	13 Capt/ 17 FO

9.3 *Apparatus*

The scenarios presented in this study were generated by a modified version of NASA's Stone Soup Simulator (SSS). The SSS is a desktop version of NASA's Advanced Concepts Flight Simulator (ACFS) which represents cockpit instruments and controls very similar to a B757 aircraft. A Silicon Graphics Indigo workstation drove the simulation that presented cockpit instruments and displays on two 19" Silicon Graphics monitors. The left monitor presented the Primary Flight Display (PFD), Navigation Display (ND), map controls, warning lights and Mode Control Panel (MCP) (see figure 5). The right monitor presented engine indications, flaps, gear, speed brakes, radio control panel, FMC CDU and a data link display. Pilots interacted with the onscreen controls via a mouse. A side stick was also available to manually fly the aircraft if required.

NASA's Stone Soup Simulator



Left Screen

Right Screen

Figure 5. Left and Right simulator display screens.

9.4 *Data Link Systems and Procedures*

The FMC CDU and MCP procedures and equipment used to control the simulator were essentially the same as those on the B757 aircraft with one major addition - all experimental conditions employed some form of data link equipment and procedures. During the course of each scenario, the participants received a set of five pre-scripted ATC data link clearances. Pilots in each of the three experimental conditions (no gating, gating/text, gating/graphic) received identical data link clearances; however, they used different data link systems and procedures to respond to those clearances. These systems and procedures were based on those employed in previous data link studies.

9.4.1 No gating condition. In all three conditions, when a data link clearance was received, an auditory alert (chime) sounded and a visual indicator flashed over the PFD (yellow for routine, red for urgent) to immediately draw attention to the clearance. In order to view the data link clearance, the pilot used the mouse to click on the "view" button below the data link screen on the left monitor (see figure 6). The "view" button was implemented in order to eliminate response variability arising from differences in the time required to recognize that a data link clearance was pending. Once the "view" button was pressed, the text of the pending ATC clearance was shown on a dedicated data link display until the pilot pushed the "view" button in response to the next data link clearance.

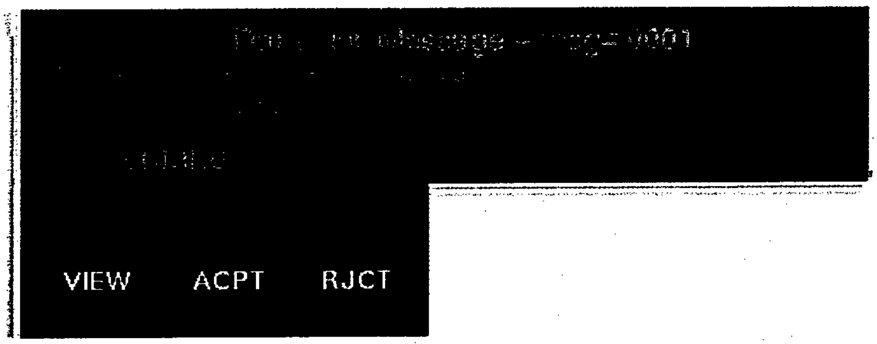


Figure 6. Data link screen in the no gating condition.

Upon viewing the data link clearance, pilots in the no gating condition were required to manually enter the appropriate targets into the FMC CDU and MCP and activate the desired autopilot and autothrottle modes. Pilots indicated clearance acceptability to ATC by clicking on the "accept" or "reject" buttons adjacent to the view button below the data link screen. In order to replicate the results of previous data link studies which examined the effects of gating on clearance loading/acceptance strategies, pilots were instructed that they could either accept the clearance first and then enter the data, or enter the data, then accept the clearance. Other than the delivery and acceptance of the pre-scripted clearances, all other air traffic control transactions (including negotiation of conflicting clearances) were conducted via voice. The experimenter played the role of air traffic controller for those voice transactions.

9.4.2 Gating/Text condition. In this condition, the alerting and initial display of the data link message was identical to the no gating condition. However, instead of manually entering clearance information, it could be loaded automatically into the FMC CDU and MCP by depressing the "load" button. Once the pilot activated the "view" button, the text of the data link clearance and a depiction of the associated performance targets appeared on a text based data link screen (see figure 7).

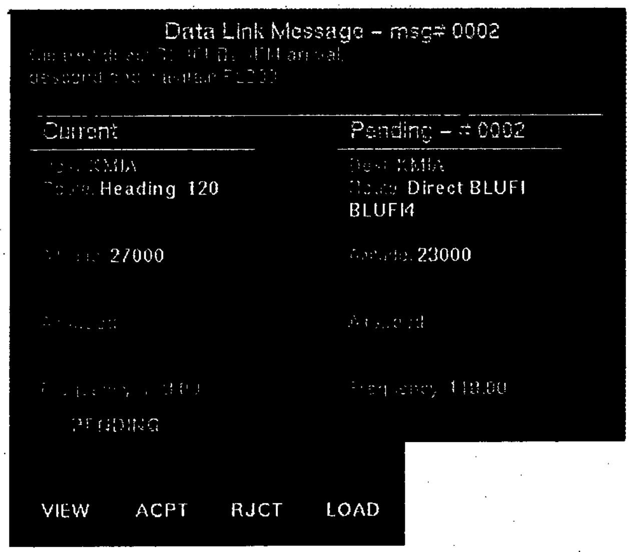


Figure 7. Data link display in the gating/text condition.

This screen was comprised of two sections - the top section was identical to the data link display used in the no gating condition and displayed the text of the current data link message. The lower section was intended to highlight for pilots the changes that would result from gating the data link clearance. This portion of the display listed the current and proposed FMC and MCP targets (routing, altitude, airspeed, etc.) in two adjacent columns. Differences between the current and proposed targets were highlighted in orange in order to draw the pilot's attention to the effects of loading and accepting a data link clearance. The text of the data link clearance remained visible until the "view" button was depressed in response to the next data link clearance. However, since the display of performance targets was intended to assist the pilot prior to the consent decision, the columns of data in the lower section of the display disappeared 10 seconds after accepting and loading (or rejecting) the clearance.

Once the pilot read the clearance, he or she could accept or reject the clearance by clicking on the "accept" or "reject" buttons below the data link display. The pilot could also automatically load the data link information directly to the FMC CDU and MCP by clicking on the "load" button. Pilots were instructed that they could load and accept the clearance in either order and that they should use the 'load" button to load clearance information (as opposed to manually entering the information) unless they rejected the clearance or found a specific conflict with the clearance implementation.

9.4.3 Gating/Graphic condition. The procedures used by this group were identical to the procedures used in the gating/text group with the exception of the display of data link information. In this condition, once the pilot pressed the "view" button, the text of the data link clearance was presented on the data link screen and the implications of gating the data link clearance were depicted on the Navigation Display (ND) and an associated Vertical Situation Display (VSD) that appeared below the ND (see figure 8). The data link display indicating the text of the data link clearance was identical to the display used in the no gating condition with the exception of the additional "load" button located underneath the data link screen. Routing and heading changes resulting from gating the data link clearance were depicted by an orange line superimposed over the already existing depiction of route and heading information on the ND. Changes to the vertical and speed profiles were indicated in orange on a static VSD located beneath the ND. 10 seconds after accepting and loading (or rejecting) the data link clearance, the orange line was removed from the ND and the VSD went blank. Both the text and graphic displays were designed to provide pilots with the same information - they both presented current and proposed targets while highlighting the resulting differences in orange. The only difference between the two displays was the format in which this information was presented. It is important to note that both the text and the graphic display depicted the impending failure of the data link system to properly load or execute the data link clearance by showing the actual performance targets that would be pursued.

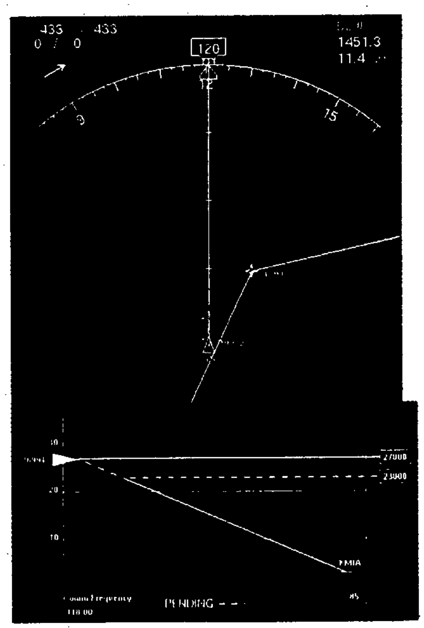


Figure 8. Data link display in the gating/graphic condition.

9.5 *Scenarios*

The eight experimental scenarios represented flight segments from near top of descent to final approach to eight different major US airports. This phase of flight was selected since the complexity, variety, and frequency of clearances given during descent allowed for a realistic, yet efficient presentation of the desired number and range of ATC clearances. During the course of each scenario, pilots received five different data link clearances. One of these five clearances was designed to present a conflict representing one of the eight previously described combinations of urgency and conflict type. One clearance in each scenario (including the conflicting clearance in half of the scenarios) was an urgent clearance; the remaining four were routine. The position of the conflicting and urgent clearances was balanced between scenarios. Table 4 indicates the general composition of the eight different scenarios. Appendix A indicates the routing and clearance information associated with each experimental scenario.

Table 4. List of scenarios and conflicts

	Level of Urgency	
Type of conflict	Routine	Urgent
Goal Conflicts		
1. Inappropriate Goal Conflict	Heading is too close to the runway (inside the final approach fix)	Speed restriction on final approach is too fast
2. Impossible Goal Conflict	Cleared to slow to an airspeed above current speed	Cleared to descend to an altitude above current altitude
Implementation Conflicts		
1. Automation does too little	A speed restriction given at cruise does not propagate to the descent phase	Failure to load an along track waypoint
2. Automation does too much	A runway change deletes a portion of the vertical profile	Deleting a previously given airspeed restriction also deletes an altitude restriction

Each subject experienced the same eight scenarios; however, the order of presentation was counterbalanced between conditions. The data link clearances used to present the eight different conflicts were designed to be as similar as possible on other relevant dimensions such as clearance length, number of targets, and conflict timing. Figure 9 indicates the general flow of one of the scenarios which represents an implementation conflict in which automated systems do more than expected.

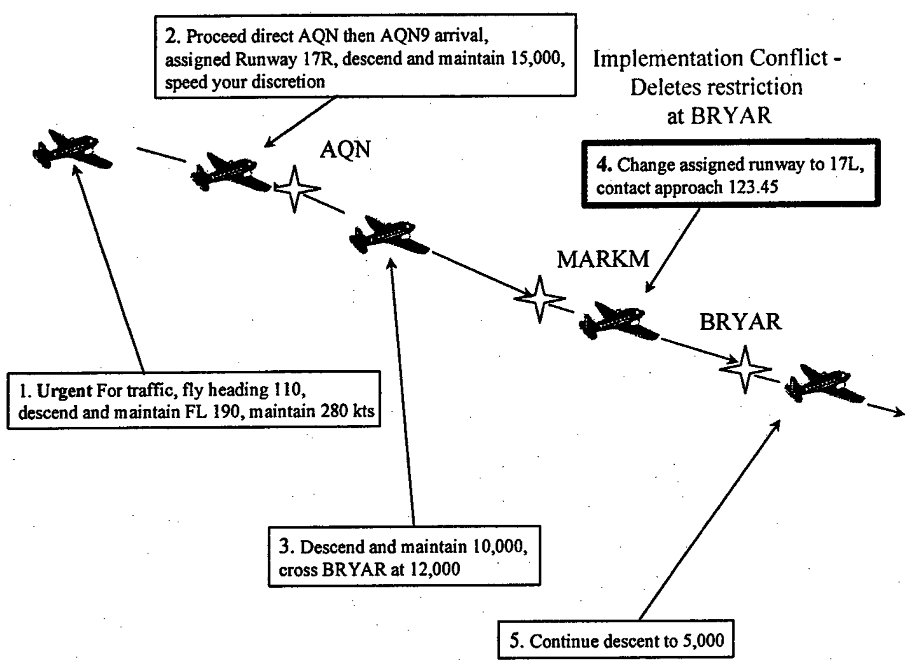


Figure 9. Example scenario showing an implementation conflict – automation does more.

9.6 *Concurrent Monitoring Task*

In order to recreate the competing attentional demands of the highly dynamic, information rich flight deck and to get some indication of pilot workload, pilots performed a concurrent monitoring task during the course of the eight scenarios. Pilots were asked to monitor a small box representing a digital load meter located above the FMC (see figure 10) for deviations above a value of 75 and were told to use the mouse to click on the display every time they noticed an out-of-limits condition. The value indicated on the digital display fluctuated randomly in the normal range and exceeded 75 on the average of once every 30 seconds. Pilots were instructed that an electrical component was malfunctioning and for maintenance purposes, they were required to keep track of the number of times the display went out of limits during the scenario. Also, pilots were instructed that, while flying the aircraft and responding to data link clearances were more critical than the monitoring task, the latter was also important and should not be ignored.

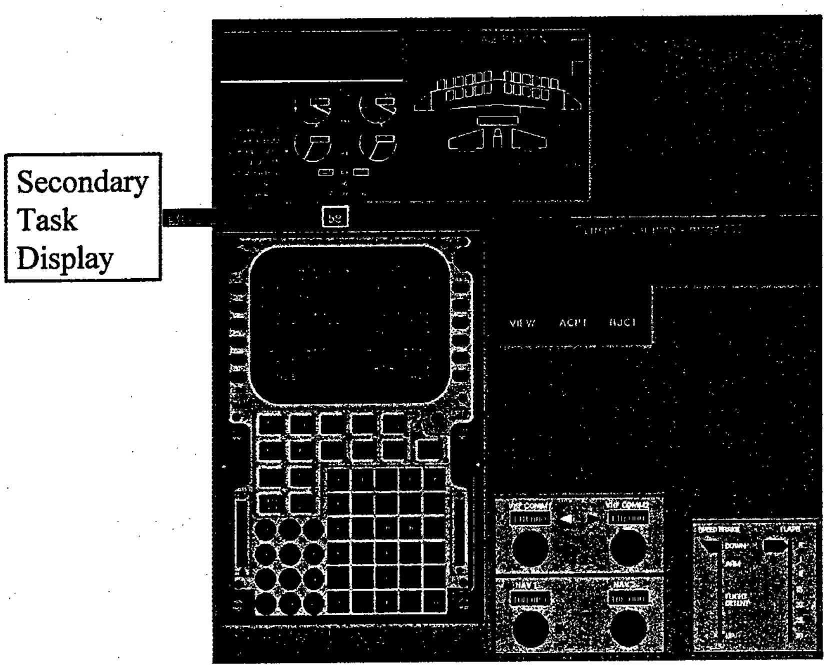


Figure 10. Secondary task display.

9.7 *Procedure*

The study was conducted in facilities provided by the participating airline. The entire experiment lasted approximately 3 ½ hours with one 10 minute break taken near the midpoint of the study. Upon arrival, participants filled out a worksheet detailing their crew position and flight experience. Participants were assigned to an experimental condition based on this demographic data. Following a brief overview of experimental purposes and procedures, pilots were given a six page set of written instructions. These instructions described the operation of the simulator and data link systems, the experimental tasks, trust ratings, and the verbal protocol data subjects would provide during the course of the experiment. The instructions indicated that the purpose of the experiment was to examine display and procedural issues associated with data link. The instructions stressed that during the course of the experiment participants should respond to ATC clearances just as they would on an actual flight. Specifically, subjects were instructed that, if a clearance seemed unclear or incorrect, they should clarify the clearance with ATC (the experimenter) or reject the clearance. Participants were also instructed that, for purposes of the experiment, they were to assume the role of the pilot flying the aircraft and would be required to make all decisions and activate all controls themselves. They were instructed that, although the autopilot would be the primary means of controlling the aircraft (as is the case in actual flight operations), manual flight control was possible through a joystick located to the right of the pilot. In order to gain concurrent verbal report data, participants were instructed to describe the questions they were asking themselves (e.g., is the proposed routing free of thunderstorms?) and the information used to answer those questions as they decided to accept, reject, or load a data link clearances. An audio tape of the verbal protocol was recorded to allow for synchronization of verbal report data and other events.

Following these instructions, participants were given a 20 minute practice scenario (a descent and instrument approach into San Diego) which presented 10 conflict-free data link clearances representative of the those used in the experimental scenarios. Subjects also performed the concurrent monitoring task and practiced providing concurrent verbal reports during this practice session. During the course of the practice scenario, pilots were encouraged to ask the experimenter for clarification regarding systems or procedures. At the end of the training session, participants were asked to provide subjective ratings of trust in the ATC controller, data link systems, as well as the FMC. The results of a preliminary pilot study indicated the duration of the practice scenario was sufficient to ensure that subjects were familiar with the operation of the simulator and data link systems.

Following the practice scenario, participants completed the eight experimental scenarios. Prior to each scenario, the participants were given an information package that briefly described the scenario including routing information, enroute and destination weather, and airfield notices (NOTAMs). An enroute map, arrival procedures, and instrument approach plates were also included in the package. After the pilot had studied this information and indicated that he/she was ready, the experimenter began the scenario. Following each scenario, participants completed the above-mentioned set of subjective ratings indicating the level of trust in ATC, data link systems, as well as the FMC.

After completion of all experimental scenarios, a debriefing session was conducted during which pilots were given more information about the purposes of the study. They were asked to provide subjective ratings detailing their desire for data link gating and describe the benefits and problems associated with both the data link system they used as well as data link gating in general. Using the verbal protocol data as well as the information accessed during the experimental scenarios, pilots were cued to describe their general strategies and problems in responding to specific data link clearances.

9.8 *Dependent Variables/Covariates*

This study collected both performance outcome and process-tracing data and also examined the relationship between these measures and subjective ratings of trust.

9.8.1 *Performance measures.* Performance measures included the speed and accuracy of operator responses to both conflicting and acceptable clearances.

Conflict detection time and accuracy. This study measured the number of detected conflicts and the latency of detection during both the consent decision (before pressing the accept and or load button) as well as during subsequent monitoring. Conflict detection time was defined as the time between pressing the "view" button and the first verbal or key press response (i.e. pressing the "reject" button or pressing the first button to begin manual data entry) indicating recognition of the conflict.

Response time. The time required to accept and load clearances as well as activate autopilot modes was measured since it may indicate the relative costs and benefits associated with gating, display design, time pressure, and the nature of the conflict. In order to provide a common reference point, all response times were measured using activation of the "view" button as the zero point. Response times to data link clearances were measured between clicking on the "view" button and 1) clicking the "load" button (pressing the "execute" button the FMC or setting a value into the MCP in the no gating condition), 2) clicking on the "accept" button, as well as 3) clicking on the required autopilot mode(s) on the MCP.

Concurrent monitoring performance. Concurrent monitoring performance may provide an indication of the cognitive effort required to extract relevant information and decide whether to accept or reject a clearance. A record of pilot responses to secondary task events (mouse clicks) was collected for each scenario.

9.8.2 *Process Tracing Measures.* Three sources of data were used to examine possible differences between cognitive processes as a result of gating, time pressure, conflict type, and display design. Verbal protocol data as well as button press data provided information concerning the type and amount of information considered during the consent decision. Button press data also indicated the timing and nature of pilot input to and interactions with the data link interface, MCP, and FMC CDU. Finally, during the debriefing, using a cued recall technique, subjects were asked to describe their decision processes across the eight scenarios.

9.8.3 *Operator Trust.* Since the perceived competence of automated systems has been determined to have the largest impact on operator trust in those systems (Muir and Moray, 1996), participants were asked to provide subjective ratings of their trust in the controller's ability to provide an acceptable clearance, trust in the ability of data link systems to translate the data link clearance into an acceptable set of performance targets, as well as trust in the ability of cockpit systems to pursue those targets in an acceptable manner. Ratings of trust were collected following each scenario (including the practice scenario) using methods adapted from Lee and Moray (1992) (see Appendix B).

10.0 **Results**

The purpose of this study was to examine the influence of gating, time pressure, display design, and trust on an operator's ability to provide informed consent in a widely distributed management-by-consent system. The effects of these factors were measured by assessing both outcome and process measures. Additionally, in order to better understand the factors contributing to the outcome and process measures, subjective data were collected during the debriefing. Prior to presenting all the results in detail, a brief summary of the major findings will be given to provide a framework for integrating the large number of dependent measures and analyses.

In general, this study found that pilots were often unable to detect goal and implementation conflicts, and thus provide informed consent prior to accepting and/or loading data link clearances. Conflict type, time pressure, gating, and trust significantly influenced detection performance, whereas display design had little effect on conflict detection performance. Goal conflicts were more likely to be detected than implementation conflicts. Within these major conflict types, impossible goal conflicts were more likely to be detected than inappropriate conflicts, and implementation conflicts in which automation did less than expected were more likely to be detected than conflicts in which automation did more than expected. Increased time pressure led to decreased detection of goal conflicts prior to consent, and in the impossible goal conflict scenarios, time pressure interacted with gating and led to reduced conflict detection only for pilots in the two gating conditions. Finally, trust in ATC had a significant relationship with conflict detection. Pilots indicating low trust in ATC and high trust in automated systems were more likely to detect conflicts prior to the consent decision; however, neither trust in

ATC nor trust in automated systems affected overall conflict detection. The three trust ratings showed significant intercorrelations.

Secondary task performance, and the response time to accept, load and execute the data link clearances were also examined. These data show that subjects in the two gating conditions detected more secondary task events, and were able to process data link clearances much quicker than subjects in the no gating condition.

An analysis of verbal protocol data as well as observed and recorded pilot actions showed that most conflicts were detected immediately before consent as pilots read and evaluated the clearance, or immediately afterwards during an initial target and performance check. Very few conflicts were detected during the subsequent monitoring stage. The analysis also indicated that gating did not eliminate errors, but instead resulted in a change in the nature of observed errors. With gating, there was a trend towards a shift from errors of commission to errors of omission.

Finally, subjective opinion data provided further explanation for the observed outcome and process results. Pilot comments showed that data link gating and the data link displays accounted for the majority of both best liked and least liked system features. In general, pilots liked the concept of data link gating. They felt that gating decreased overall pilot workload, but also created additional monitoring workload and could lead to problems with complacency. The display of the data link clearance was considered a valuable memory aid during confirmation and monitoring processes; however, it sometimes did not display applicable ATC instructions to pilots, and was located too far from other cockpit displays of aircraft status and performance.

The following sections will detail the observed outcome, process and subjective data. The first three sections will describe observed conflict detection effects and discuss a model of conflict detection that can help to explain the observed results. The following three sections will describe some of the costs and benefits of data link gating including secondary task performance, response time measures, and error data. The final section will briefly discuss subjective data collected during the debriefing.

10.1 *Conflict Detection*

In accordance with the categorical nature of conflict detection data, detection percentages were examined using Logistic Regression analyses as well as Chi-square analyses. The logistic regression analysis (Agresti, 1996) describes the relationship between an independent variable and a binary dependent measure as expressed by the probability of membership in one binary category associated with a given value of the independent variable. Logistic regression analyses can either be interpreted as a linear approximation of the logistic regression curve or as an odds ratio (the relative odds of the independent variable predicting a given value on the dependent measure). The analyses reported in this section will use the latter (odds ratio) interpretation. In other words, a statistically significant effect implies that conflict detection was more likely for one or more levels of a given independent variable compared to other levels of that variable. The results reported in this section were derived from a statistical model that considers the effects of gating/display condition, scenario, trust, and experience. The model does not include interactions between these factors due to sample size limitations. The logistic regression analyses reported in this study were conducted using SUDAAN software which considers the repeated measurement of conflict detection performance.

10.1.1 *Conflict detection before consent.* In general, conflict detection before consent was poor with less than 50% of conflicts detected in all cases (see figure 11). A logistic regression analysis indicates that goal conflicts were significantly more likely to be detected than implementation conflicts ($F(1,22) = 394.37$, $p<.0001$). Some goal conflicts (42.5% - no gating, 27.5% for both gating/text and gating/graphic conditions), but no implementation conflicts, were detected prior to the consent decision.

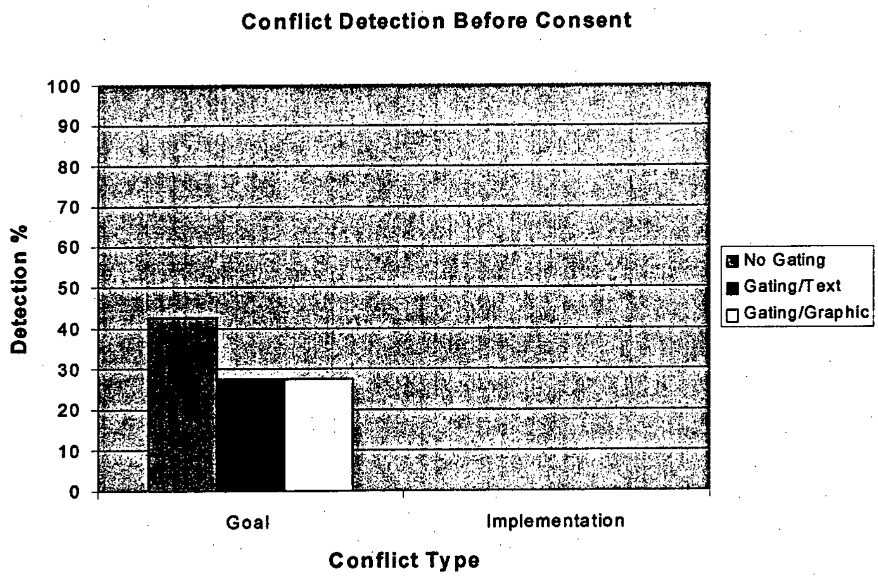


Figure 11. Conflict detection before consent – goal vs. implementation conflicts.

Figure 12 shows detection of goal conflicts before the consent decision, broken down by scenario and gating/display condition. A logistic regression analysis showed a highly significant main effect for scenario (reflecting detection differences between the combinations of conflict type and time pressure presented in the experimental scenarios) ($F(7,22) = 127.11$, $p <.0001$), and marginally significant effects due to time pressure ($F(1,22) = 3.94$, $p = .0568$) and gating/display condition ($F(2,22) = 2.91$, $p = .0706$). A closer inspection of figure 12 shows that the marginally significant effect of gating/display condition arises primarily from an interaction between time pressure and gating condition for the detection of urgent impossible conflicts.

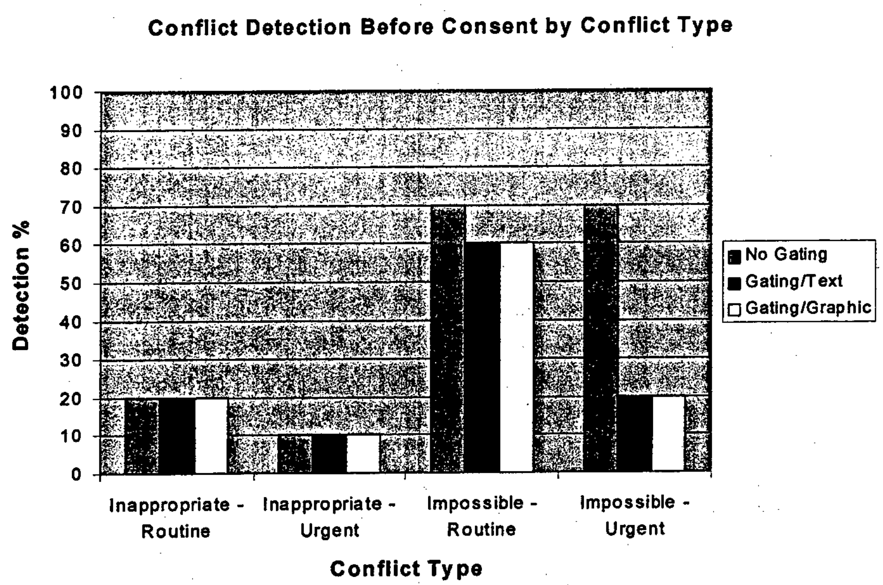


Figure 12. Detection of goal conflicts prior to consent.

Separate Chi-square analyses (see table 5 for frequency data) show that in the case of urgent clearances, subjects in the two gating conditions detected significantly fewer impossible conflicts than subjects in the no gating condition (X^2 (1, N = 30) = 7.18, p = .0074); however, there was no significant difference in detection performance between gating and no gating conditions for impossible conflicts embedded in routine clearances (X^2 (1, N = 30) = 0.29, p = .5921).

Table 5. Detection of impossible conflicts in routine and urgent clearances

	Routine		Urgent	
Gating Condition	Yes	No	Yes	No
No Gating	7 (70%)	3 (30%)	7 (70%)	3 (30%)
Gating	12 (60%)	8 (40%)	4 (20%)	16 (80%)
Total	19	11	11	19

The data were also analyzed for differences in the detection of inappropriate and impossible goal conflicts. Collapsed across experimental condition and time pressure, inappropriate conflicts (50% detected) were significantly less likely to be detected before consent than impossible conflicts (85% detected) ($F(1,22) = 13.72$, $p = .0009$).

10.1.2 *Overall conflict detection(both before and after the consent decision).* Even if conflicts are not detected prior to the consent decision, they may still be detected and corrected during subsequent confirmation and monitoring processes. Figure 13 shows the overall detection of goal and implementation conflicts (the shaded area represents conflicts detected prior to the consent decision, while the unshaded region reflects those conflicts detected following the consent decision).

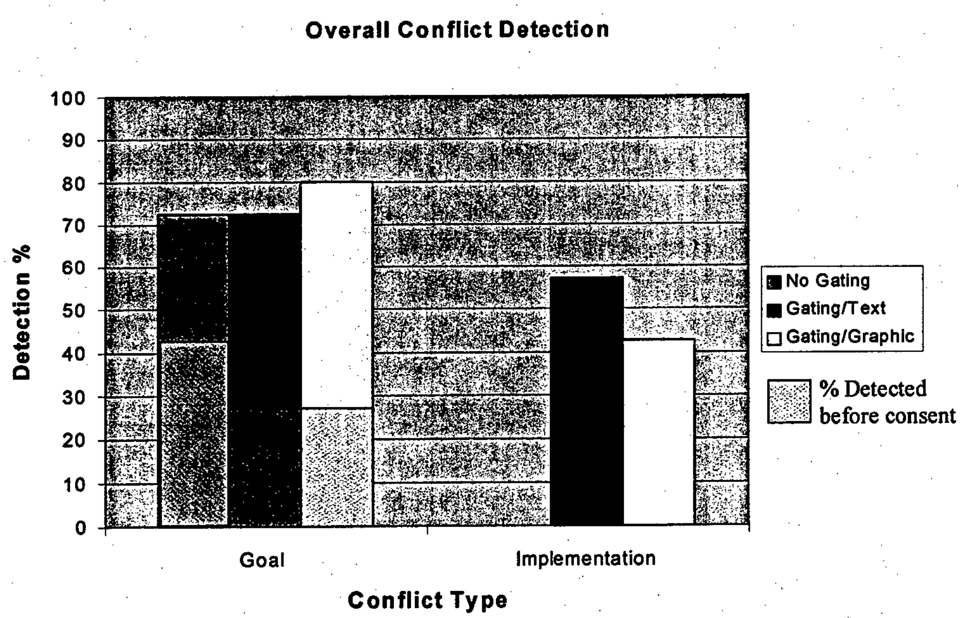


Figure 13. Overall conflict detection – goal vs. implementation conflicts.

Once again, significantly more goal conflicts (75%) were detected than implementation conflicts (49%) ($F(1,22) = 4.10$, $p = .0522$). Note that pilots in the no gating condition could not respond to implementation conflicts since these conflicts were created by the automatic loading process employed only in the two gating conditions.

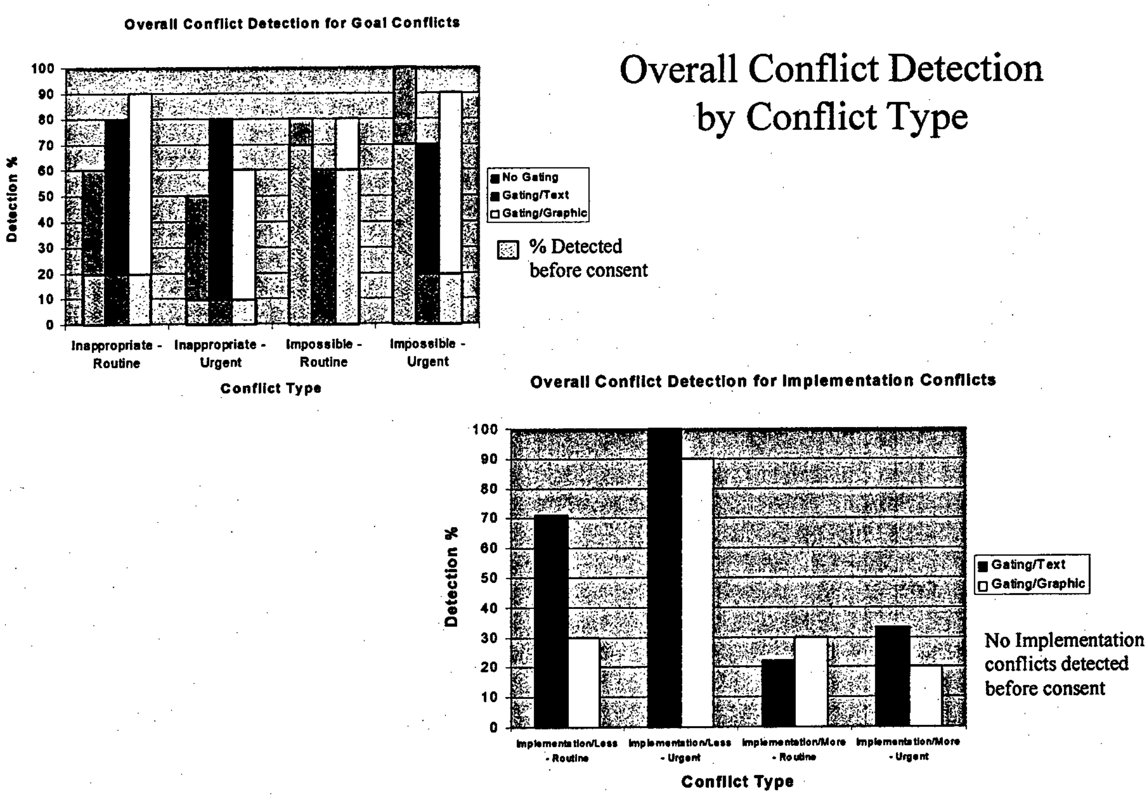


Figure 14. Overall conflict detection broken out by conflict type.

Figure 14 shows overall conflict detection broken down by scenario and display/gating condition. A logistic regression analysis revealed a highly significant main effect for scenario ($F(7,22) = 7.86$, $p < .0001$), but no significant effect due to gating/display condition ($F(2,22) = 0.65$, $p = .5277$). Further comparisons provide a more complete picture of the effects of conflict type on overall conflict detection.

Implementation conflicts in which automation did less than expected were significantly more likely to be detected than conflicts in which automation did more than expected ($F(1,22) = 17.26$, $p = .0003$). Also, while impossible goal conflicts were significantly more likely to be detected than inappropriate goal conflicts *prior to the consent decision*, there was a non-significant difference in *overall detection* of impossible and inappropriate conflicts ($F(1,22) = 2.14$, $p = .1546$) with a trend toward greater likelihood of detecting impossible conflicts.

In order to compare these results to previous findings by Hahn and Hansman (1992), Chi-square analyses were conducted to assess the effects of gating on overall detection of inappropriate conflicts. Due to required independence assumptions of the Chi-square test, detection performance could not be collapsed across routine and urgent conditions for purposes of this analysis (see table 6 for detection frequencies). The results show a non significant trend towards superior detection of inappropriate conflicts by pilots in the gating conditions for routine clearances (X^2 (1, N = 30) = 2.33, p = .1270 and urgent clearances (X^2 (1, N = 30) = 1.15, p = .2839).

Table 6. Overall detection of inappropriate conflicts in routine and urgent clearances

	Routine		Urgent	
Gating Condition	Yes	No	Yes	No
No Gating	6 (60%)	4 (40%)	5 (50%)	5 (50%)
Gating	17 (85%)	3 (15%)	14 (70%)	6 (30%)
Total	23	7	19	11

10.2 *Processes Involved in Conflict Detection*

Outcome measures such as conflict detection provide useful indications of the overall performance effects of gating, display design, time pressure and trust. In order to achieve a more complete understanding of how these effects are brought about and what stages of information processing are primarily affected, process data are also needed. Information derived from verbal protocol data, a record of pilot button presses, and observations made by the experimenter during the course of the experimental scenarios were combined with an analysis of pilot tasks to derive a model of pilot conflict detection processes during and after acceptance and loading of a data link clearance. The following sections explain this model, present an analysis of the relative effectiveness of various cognitive processes across the set of independent measures employed in this study, and describe the time course of these conflict detection processes.

10.2.1 *An error detection model.* Similar to Sellen's (1994) model of human error detection, our model indicates the stages and processes that could lead to the detection of goal and implementation conflicts (see table 7). The left hand columns identify the stages and observable actions exhibited by pilots, while the right hand column indicates the associated cognitive conflict detection processes. This model proposes four stages, one stage prior to the consent decision called pre-consent evaluation, and three stages following consent - post-consent confirmation, subsequent monitoring, and detection via forcing function. Note that while the post-consent confirmation and subsequent monitoring stages share similar underlying processes, they fulfill distinctly different purposes. The post-consent confirmation stage serves to ensure that desired performance targets have been entered and that the *initial* aircraft performance is proceeding as expected (e.g. the aircraft was beginning to climb or turn). In contrast, during the subsequent monitoring stage, indications of aircraft performance are evaluated to ensure that the aircraft performance is meeting specific parameters (e.g. climbing to a specific altitude or rolling out on a specific heading). Both expectation-driven and data-driven monitoring processes were observed during that stage. Expectation-driven monitoring was generally preceded by a rereading or verbalization of the ATC clearance followed by a scan of specific displays to confirm that aircraft behavior matched the ATC clearance. Data-driven conflict detection occurred during a general scan of cockpit instruments not specifically related to the requirements of the ATC clearance.

Table 7. A model of conflict detection stages and processes

Stage	Observable Actions	Cognitive processes
Pre-consent Evaluation	• Read entire clearance/portion of clearance • Load portion of clearance (no gating only)	a) Read and understand clearance b) Activate knowledge of aircraft state c) Activate knowledge of flight procedures and regulations associated with current context d) Compare and assess compatibility of clearance with b) and c) above
Post-consent Confirmation	• Reread entire clearance • Confirm data entry/mode selection • Confirm initial aircraft performance	a) Reread clearance b) Identify associated performance targets and expected initial aircraft behavior c) Search for targets in MCP and FMC CDU, search for indications of initial changes in aircraft behavior on PFD and ND d) Compare b) and c) to identify discrepancies
Subsequent Monitoring	• Expectation-driven - Reread clearance then scan instruments to confirm • Data-driven – general instrument scan leads to observations of discrepant behavior and further information gathering	a) Recall/reread desired performance parameters b) Observe aircraft performance on ND and PFD c) Compare a) and b) d) If discrepancy is noted, develop initial diagnosis e) Search for indications to support hypothesis
Forcing Function/ Outside Intervention	• Other cockpit tasks (e.g. the need to slow to deploy landing flaps) force recognition of conflict • ATC (experimenter) intervenes to keep performance within scenario limits	a) Identify concurrent and projected tasks b) Determine requirements of those tasks c) Search PFD, ND, MCP, FMC for indications of current performance an modes d) Compare b) and c) for discrepancies

10.2.2 *Conflict detection effectiveness at different stages.* This conflict detection model was used to categorize and analyze all conflict detection episodes observed in this study. This analysis was conducted by first identifying the point at which the initial indication of conflict recognition was noted. The experimenter then reviewed the audio tape of the verbal protocol provided by the pilot as well as the record of button presses and notes made during the experimental session in order to determine the activities and processes involved in detection of a particular conflict. For example, if a pilot was observed to deploy the flaps for landing (as confirmed afterward by control activation data), just prior to stating: "Wait a minute, he wanted 220 knots", this episode of conflict detection was categorized as detection resulting from a forcing function in which the demands of another cockpit task (the need to slow down to lower the flaps) were evaluated against the current ATC clearance (maintain 220 knots) thus triggering conflict detection.

Based on these data, the processes that resulted in error detection were identified and categorized. Table 8 lists the frequency of goal and implementation conflicts detected at each stage (see Appendix C for a complete breakdown). The majority of all conflicts were detected during the pre-consent evaluation (31.2%) and post-consent confirmation stages (44.0%). Relatively few conflicts were detected during the subsequent monitoring processes (18.4%). Pre-consent evaluation and forcing functions were effective in aiding detection of goal conflicts only.

Table 8. Frequency of conflicts detected during each stage by gating/display condition and type of conflict

Conflict Detection Stage	No Gating Goal	Gating./Text		Gating/Graphic		Total
		Goal	Implement -ation	Goal	Implement-ation	
Pre-Consent Evaluation	17	11	0	11	0	39 (31.2%)
Post-Consent Confirmation	6	9	16	13	12	55 (44.0%)
Subsequent Monitoring	4	6	4	4	5	23 (18.4%)
Forcing Function	2	3	0	3	0	8 (6.4%)
Total	29	29	20	32	17	125

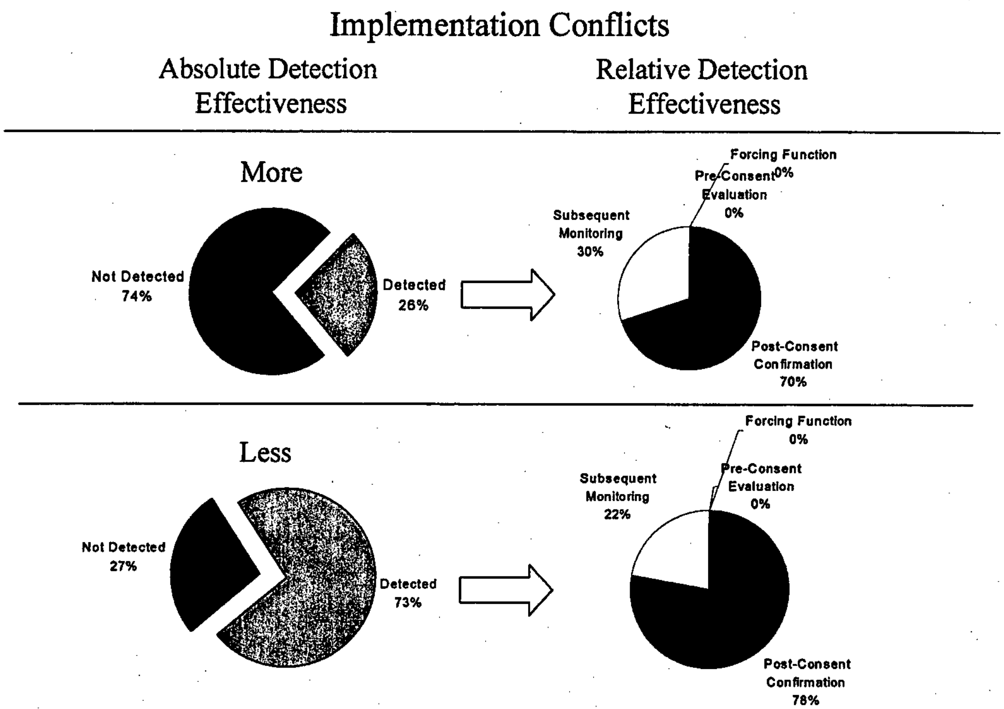
In addition to identifying conflict detection stages, this analysis also examined the conflict detection triggers employed at each stage (see table 9). Conflict detection during the <u>pre-consent evaluation</u> stage was triggered primarily by a comparison between the <u>requirements of other tasks and procedures</u> and the instructions contained in the data link clearance (e.g. comparison between the requirement to intercept the final approach course outside the final approach fix vs a heading contained in the clearance). During the <u>post-consent confirmation</u> stage, most conflicts were detected by a comparison between the instructions specified in the clearance and reference to <u>targets in the MCP and FMC</u>. During this stage, only *goal conflicts* were detected by reference to the text of the data link clearance, whereas most *implementation conflicts* (24/28) were detected by reference to data entered into the FMC and MCP. During the <u>subsequent monitoring</u> stage, detection was most often triggered by a comparison between desired and <u>actual aircraft behavior</u>; however, some detection occurred while rereading the data link clearance prior to a scan of cockpit instruments.

Table 9. Conflict detection triggers broken down by conflict detection stage

Conflict Detection stage	Trigger: Requirements of clearance vs Displays of current aircraft performance	Trigger: Text of the data link clearance vs Requirements of other tasks/procedures	Trigger: Targets in clearance vs Targets in FMC/MCP
Pre-consent Evaluation	9	**30**	0
Post-Consent Confirmation	19	12	**24**
Subsequent Monitoring	**12**	8	3
Total	40	50	27

This analysis of conflict detection stages and triggers informs the design of future displays by identifying the elements of information used to detect conflicts at each stage and can help us better understand the mechanisms by which factors such as conflict type and time pressure affect conflict detection performance. The following paragraphs will examine differences in conflict detection stages and triggers between: 1) implementation conflicts in which automation does more and automation does less, and 2) routine and urgent impossible conflicts.

Implementation conflicts – Automation does more vs automation does less. Conflicts in which automation did more than expected were detected less often than conflicts in which automation did less. Figure 15 shows the stages at which these two types of implementation conflicts were detected. As indicated on the right side of this figure, there was little difference in the relative effectiveness of the post-consent confirmation and subsequent monitoring stages. These results indicate a *general decrease* in detection effectiveness across stages rather than *specific difficulties* in either one of these



stages.

Figure 15. Conflict detection processes for scenarios were automation did more vs. automation did less.

The effects of time pressure. Time pressure had a large impact on the detection of impossible conflicts only in the two gating conditions. As shown in the top portion of figure 16, time pressure had little effect on conflict detection in the no gating condition. The bottom portion of this figure indicates that, for pilots in the two gating conditions, time pressure led to reduced conflict detection effectiveness at the pre-consent evaluation stage. Further analysis showed that, under time pressure, detection of impossible conflicts during the post load evaluation stage was triggered exclusively by a discrepancy between expected and observed aircraft performance.

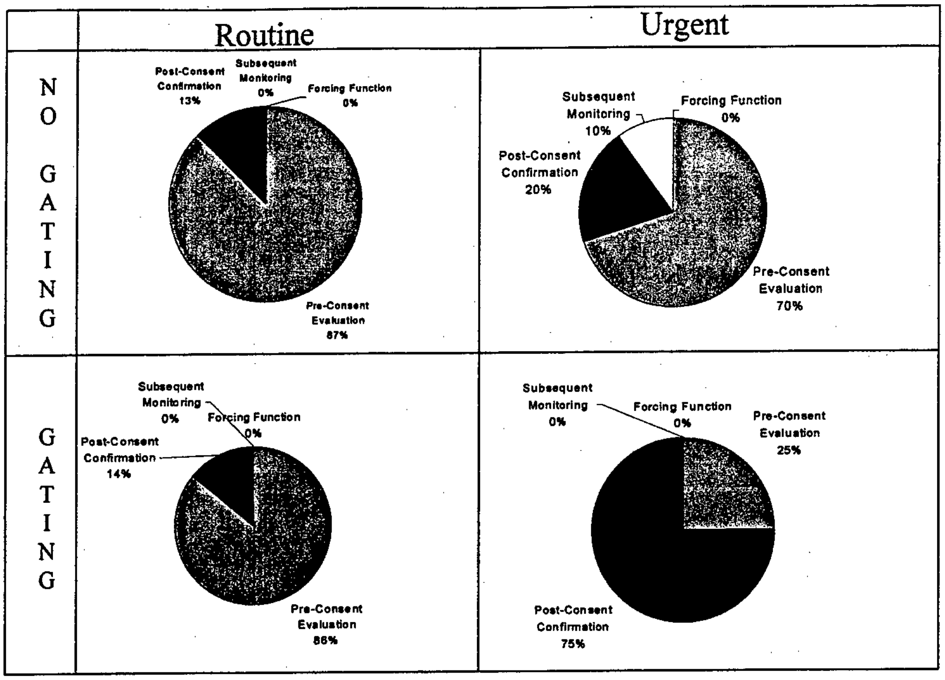


Figure 16. Conflict detection processes for routine vs. urgent impossible goal conflicts.

Decreased depth of processing under time pressure may account for the relative inability of pilots in the gating conditions to detect impossible conflicts during the pre-consent evaluation stage. Both performance and observational evidence support this interpretation. First, the average "accept" time for subjects in the two gating conditions who did not detect the conflict prior to the consent decision was 7.21 seconds (S.D. = 1.97) for the routine impossible conflict, and 4.70 seconds (S.D. = 1.23) for the urgent impossible conflict. Although caution must be used in interpreting these results due to the relatively small sample sizes, t- tests show that the "accept" time in the urgent condition was significantly faster than in the routine condition for both gating conditions (t = 4.09 (df = 25), p = .0002). This suggests that under high time pressure, subjects in the two gating conditions spent considerably less time evaluating the acceptability of a clearance prior to the consent decision. Note that the response time of 4.7 seconds was also considerably shorter than the time available (15 seconds was the limit for an urgent clearance).

Under time pressure, some subjects were observed to read only part of the pending data link clearance. The verbal protocol data also indicate that, while the separate portions of the clearance were read and understood locally, pilots did not evaluate the clearance for conflicts at a global level. This process is typified by one of the subjects in the gating/text condition who read, accepted, and loaded a clearance containing an impossible conflict which instructed the pilot to descend to an altitude above the current altitude. After loading the clearance, the subject checked to confirm that the desired altitude was loaded and activated the desired vertical mode. The pilot noted that something was wrong when the aircraft began climbing rather than descending. Puzzled by this behavior, the pilot reread the clearance and checked that the correct altitude had been loaded into the MCP. The pilot then began to explore the possibility that the selected vertical mode was malfunctioning before finally contacting ATC for clarification. This example again illustrates that time pressure led to decreased depth of processing.

10.2.3 *Time course of conflict detection stages.* In addition to the relative effectiveness of the conflict detection processes, the time course of conflict detection is also a concern. Figure 17 indicates the average conflict detection time across all scenarios for subjects in each gating/display condition.

Since conflict detection can occur at several discrete stages, the underlying distribution of detection times is multi-modal and difficult to analyze and interpret. For example, the high average detection time for the "urgent inappropriate" conflicts is not due to a general increase in detection time across all stages. Instead, this difference is primarily due to the fact that a large proportion (25%) of these conflicts were detected due to forcing functions which tend to take effect only after considerable delay.

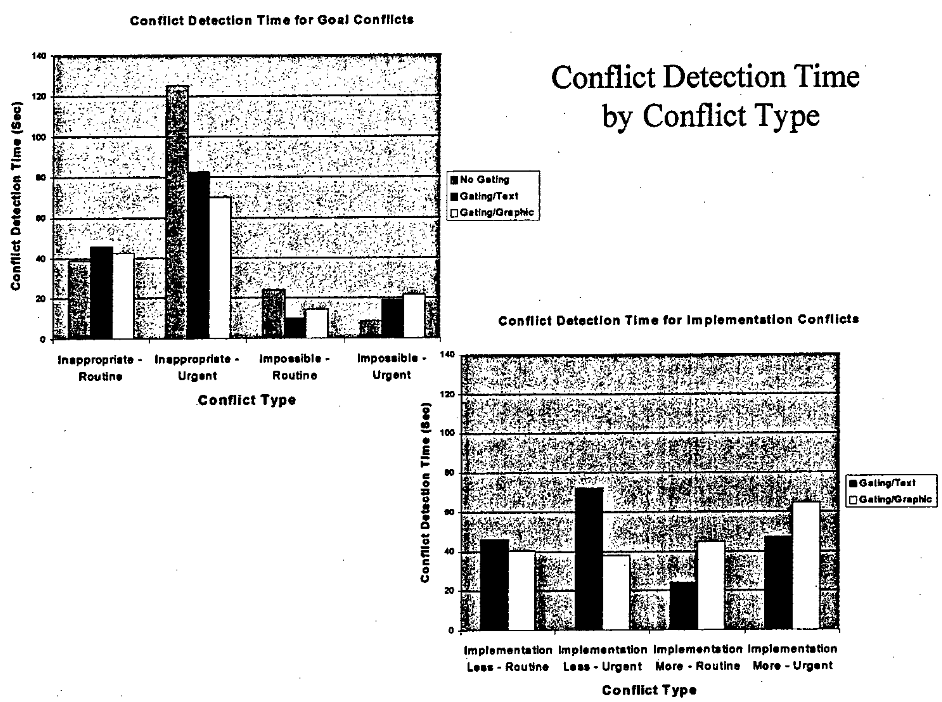


Figure 17. Average conflict detection time across scenarios by gating/display condition.

Therefore, instead of analyzing detection time differences due to conflict type, display condition, and time pressure, we will examine the average conflict detection time for each conflict detection stage (see table 10). Due to the small number of observations in some cells, as well as the skewedness of the underlying distributions, statistical analyses were not conducted on these data. It appears, however, that if conflicts are not detected during the pre-consent evaluation or post-consent confirmation stages, detection tends to be substantially delayed.

Table 10. Average detection time (in seconds) for various conflict detection stages broken down by gating/display condition

Conflict Detection Stage	No Gating	Gating/ Text	Gating/ Graphic
Pre-Consent Evaluation	15.39	9.249	10.42
Post-consent confirmation	50.28	36.84	28.84
Subsequent Monitoring	56.55	90.45	61.45
Forcing function	238.35	134.62	124.56

10.3 *Trust and Experience*

10.3.1 *Trust.* Three separate subjective trust ratings were collected following each scenario – trust in ATC, trust in the data link systems, and trust in the FMC. The following subsections will examine changes in trust across scenarios, the correlation between trust measures, and the relationship between trust and conflict detection.

Changes in trust. Pilots provided subjective trust ratings on a 1 – 7 scale immediately following each scenario, with 1 corresponding to low trust and 7 corresponding to high trust. Figure 18 shows the initial (following the practice scenario) and final (immediately after the last experimental scenario) median trust ratings for each gating/display group. Note that for the no gating group, only trust in ATC and the FMC are reported since the no gating group had no experiences on which to rate trust in data link.

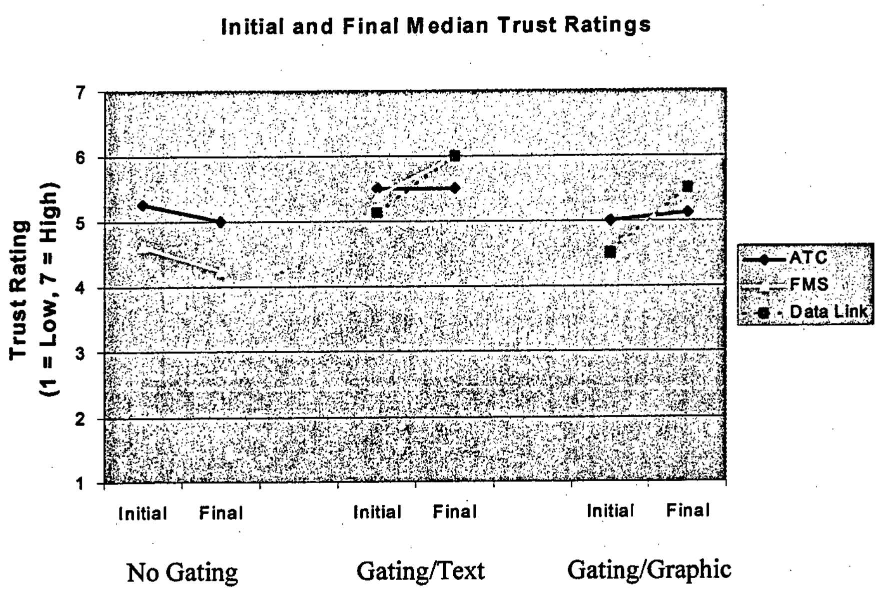


Figure 18. Change in trust – Initial vs. final median trust ratings.

In general, pilots indicated moderate levels of trust in all three areas as evidenced by median ratings above the neutral point (4). In order to examine changes in trust over the course of the study, separate Wilcoxon signed ranks tests were performed (see table 11). In both gating conditions, trust in data link systems and the FMC showed a significant increase over the course of the experiment while trust in ATC remained unchanged. For the no gating group, there were no significant changes for either trust in ATC or trust in the FMC.

Table 11. Wilcoxon Sign test results for significance of change between initial and final trust measures

	Trust in								
	ATC			Data Link			FMC		
Condition	T	*n*	*p*	T	*n*	*p*	T	*n*	*p*
No Gating	4	5	>.5				8	6	>.5
Gating/Text	9.5	6	>.5	32	6	**.008**	36.5	9	**.056**
Gating/Graphic	15	7	.47	48	10	**.019**	51.5	10	**.006**

While trust generally increased over the *course of the entire experiment*, trust sometimes decreased temporarily as a result of detected goal and implementation conflicts in a particular scenario. Sixty - seventy percent of all subjects frequently indicated changes in trust across scenarios. Since conflicts were presented in a counterbalanced order, it is impossible to analyze the time course of trust decrement and recovery; however, a general analysis of changes in trust may provide some indication of the relationship between conflict detection and subjective trust. Table 12 indicates the frequency of trust increases and decreases over the course of those scenarios in which a conflict was detected. These results show that trust most often remained constant or increased over the course of a scenario. This general trend was observed across scenarios. In all scenarios, a majority of pilots indicated no change or an increase in trust in automated systems. In only one scenario - the urgent impossible goal conflict - did a majority (60%) of pilots indicate a decrease in trust ATC.

Table 12. Relative change in trust for scenarios in which a conflict was detected (n=127)

	Change in Trust		
Trust in:	Decrease	No Change	Increase
ATC	33.1%	43.3%	23.6%
Data Link	19.7%	48.0%	32.3%
FMC	20.5%	52.0%	27.6%

The interrelationship between trust measures. In order to determine the extent to which operators can apportion trust between human and machine agents in the overall system, Pearson correlation coefficients (controlling for the effects of scenario and subject) were computed for the three trust measures as well as for changes in trust for those scenarios in which a conflict occurred (see tables 13 and 14 respectively). These analyses shows a strong positive relationship between data link and FMC trust ratings and a somewhat weaker positive relationships between ATC and data link/FMC trust measures.

Table 13. Correlation between trust measures across all scenarios (n = 235)

	Trust ATC	Trust DL	Trust FMC
Trust ATC	1.0		
Trust DL	**.342***	1.0	
Trust FMC	**.307***	**.849***	1.0

* p <.001

Table 14. Correlation between *changes* in trust measures across those scenarios in which a conflict was detected (n = 94)

		Change in:	
	Trust ATC	Trust DL	Trust FMC
Change in trust - ATC	1.0		
Change in trust - DL	**.192 ***	1.0	
Change in trust - FMC	-.007	**.440****	1.0

* p=.07, ** p <.0001

The effects of trust on conflict detection. Logistic regression analyses were used to examine the effects of trust (as measured by trust prior to each scenario) on conflict detection performance. Since trust in data link and the FMC were highly interrelated, these two trust measures were combined in an additive manner into a single measure – trust in automated systems. Separate analyses were conducted to examine the effects of trust on detection before the consent decision as well as overall conflict detection. Both trust in ATC and trust in automated systems were broken down into four categories in order to account for the possibility of a nonlinear relationship (see table 15). Trust data were categorized using category frequency and natural breakpoints in the trust distribution as criterion. With respect to conflict detection before consent, there was a marginally significant effect of both trust in ATC ($F(3,22) = 2.86$, $p = .0540$) and trust in automated systems ($F(3,22) = 2.59$, $p = .0720$). There was no effect of trust on overall conflict detection (trust in ATC - $F(3,22) = 1.38$, $p = .2694$; trust in automated systems - $F(3,22) = 0.12$, $p = .9489$).

Table 15. Odds ratios for conflict detection before consent for trust in ATC and automated systems

Measure	Odds Ratio	95% Confidence Interval
Trust in ATC (1 = low, 7 = high)		
1.0 – 3.99	90.92	2.27 – 3648.42
4.0 – 4.99	10.26	1.19 – 88.14
5.0 – 5.99	1.08	0.22 – 5.23
6.0 – 7.0	1.00	
Trust in Automation (2 = Low, 14 = high)		
2.0 – 7.99	0.12	0.00 – 2.86
8.0 – 9.99	0.08	0.01 – 1.00
10.0 – 11.99	0.10	0.01 – 0.88
12.0 – 14.0	1.00	

In order to better understand the effects of trust on conflict detection before consent, the odds ratios generated by these analyses were examined (see table 15). The depicted odds ratios indicate the likelihood of conflict detection by pilots in a given category relative to pilots in a reference category (the bottom category) which is assigned an odds ratio of 1.00. For example, when pilots rated trust in ATC from 4.0 – 4.9, they were 10.26 times as likely to detect conflicts than when they indicated higher trust ratings from 6.0 – 7.0. The associated confidence interval provides an assessment of the reliability of that estimate. Two groups have significantly different odds ratios if the odds ratio of one group falls outside the confidence interval associated with the other group. An examination of the depicted odds ratios suggestss that when pilots indicated lower trust in ATC (< 5.0) and extremely high trust in automated systems (>12.0), they were most likely to detect conflicts prior to consent. Note that the findings with respect to trust in automated systems must be interpreted cautiously due to some overlap with the 95% confidence intervals.

10.3.2 *Experience.* Logistic regression analyses were used to assess the effects of experience on conflict detection. Similar to trust data, experience data were categorized on the basis of minimum category size and natural breakpoints (see table 16). Since crew position was highly correlated with total flight experience ($r = .664$), it was not included in the logistic regression model. Both total flight experience and experience in the B757 had a significant effect on detection before consent (total hours $F(3,22) = 8.06$, $p = .0005$; B757 hours $F(3,22) = 3.69$, $p = .0229$) as well as overall conflict detection (total hours $F(3,22) = 4.40$, $p = .0114$; B757 hours $F(3,22) = 3.59$, $p = .0253$).

Table 16. Odds ratios for flight experience measures

Measure	Detection Before Consent		Overall Detection	
	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval
Hours in B757				
0 - 600	0.44	0.05 – 3.69	0.37	0.09 - 1.57
600 - 1500	0.07	0.01 – 0.65	0.16	0.04 - 0.75
1500 - 2500	1.22	0.27 – 5.61	1.11	0.38 - 3.26
>2500	1.00		1.00	
Total Hours				
0 - 5000	9.19	2.20 – 38.42	5.20	1.48 - 18.29
5000 - 10000	24.76	5.00 – 122.53	2.43	0.96 - 6.11
10000 - 18000	0.28	0.02 – 4.33	0.48	0.13 - 1.72
>18,000	1.00		1.00	

Table 16 lists the odds ratios generated by this analysis. These results show that conflicts were *least* likely to be detected (both overall and before consent) by pilots with 600 – 1500 hours in the B757, and by pilots with greater overall flight experience (> 10,000 hours).

10.4 *Secondary Task Performance*

In order to examine the effects of DL gating on the performance of concurrent tasks, this study employed a secondary monitoring task. For purposes of this analysis, event detection was defined as an indication of detection within 15 seconds of event onset. This definition was adopted in order to eliminate false alarms while still allowing for somewhat delayed responses.

In order to compensate for variations in the number of possible events, secondary task performance was computed as the percentage of events detected out of the total number of possible events. Table 17 indicates average event detection percentage and standard error for subjects in each condition.

Table 17. Secondary task performance - % of events detected

	Detection	
Condition	%	Std Dev
No Gating	16.2	13.0
Gating/Text	25.2	16.5
Gating/Graphic	23.0	16.8

These results were analyzed via a two way Analysis of Variance (ANOVA) using gating/display condition and scenario as fixed factors and subject as a random factor. This analysis showed a non- significant main effect for gating/display condition ($F(2,27) = 1.96$, $p = .1601$) with a trend towards superior performance in the two gating conditions. There was no effect for scenario ($F(7, 182) = 1.25$, $p = .2792$), and no condition x scenario interaction was observed ($F(14, 182) = .70$, $p = .7712$).

10.5 *Response Time Measures*

The time required to accept, load, and select the modes associated with an incoming data link clearance provided some measure of the potential costs and benefits of data link gating and display design. In contrast to the conflict detection time data discussed in section 10.3.3, the response time data discussed in this section were collected in response to clearances that did not contain an experimenter–induced conflict. Button press data were collected from the computer simulation at a rate of 30 Hz. The accept, load, and mode selection times were measured from activation of the "view" button until activation of the button of interest. Table 18 indicates the time required to accept, load, and activate the corresponding autopilot modes. These results show a general speed advantage for subjects in the two gating conditions. The only exception to this general trend is for activation of the heading select mode, which was slightly faster for pilots in the no gating condition. Note that response times for pilots in the no gating condition are split into two categories – those who "accepted" the clearance prior to manually loading the data (4/10 pilots), and those who loaded and executed the data first before "accepting" the clearance (6/10 pilots).

Table 18. Response times broken down by gating/display condition (sec)

Condition	Accept	Load	HDG SEL	LNAV	Vertical
No gating Load /Accept	40.52	32.26	17.2	26.77	26.01
No Gating Accept/Load	9.31	37.16	13.92	35.33	29.5
Gating Text	8.30	9.79	21.65	17.99	17.84
Gating Graphic	6.65	8.23	19.67	16.24	15.67

10.6 *Pilot Errors*

Data link gating was developed, in part, to reduce errors related to manual data entry. However, as indicated by previous research (Sarter and Woods, 1997), the introduction of automated systems often creates the opportunity for new types of errors instead. For example, it can lead to a tradeoff between a decrease in errors of commission (such as data entry errors) and an increase in errors of omission (failures to engage modes or make other control entries). In order to assess the relative frequency of these error types, responses to non-conflicting clearances were examined for data entry and mode activation errors. For purposes of this analysis an error was defined as either 1) entering data that did not match the current data link clearance, or 2) failure to activate the mode required to pursue an intended target in the MCP or FMC CDU (e.g. failure to engage a vertical mode to initiate a descent). Furthermore, this analysis only included those instances in which the errors went undetected by the pilot, thus forcing the experimenter to intervene to keep the simulation within the limits of the planned scenario. This definition of error is rather conservative since it eliminates from consideration those instances in which data entry errors or mode selections were detected and corrected by the pilot in time to prevent a problem.

Table 19. Observed undetected errors not due to experimenter-induced conflicts

	Gating/Display Condition		
Error	No Gating	Gating/Text	Gating/Graphic
Failure to press "accept"	15	0	0
Failure to press "load"	0	1	0
Misload data into the FMC	12	0	0
Failure to engage Heading Select	3	6	5
Misload data into the MCP	4	2	0
Portion of ATC Clearance ignored	0	2	2
Failure to engage LNAV	2	1	1
Failure to engage a vertical mode	0	0	2
Total	36	12	10

Table 19 lists the types of uncorrected data entry and mode activation errors committed by pilots broken down by gating/display condition. This analysis shows that pilots in the no-gating group frequently forgot to activate the "accept" button to indicate clearance acceptance to ATC, and also made many data entry errors. However, as evidenced by the number of failures to activate autopilot modes, gating did not eliminate errors altogether. Mode activation errors were committed by pilots in the all conditions. In general, errors in the two gating conditions were almost exclusively (20/22) errors of omission, while almost half (16/36) of the errors in the no gating condition were errors of commission.

10.7 *Subjective Questionnaire Data*

During the debriefing, pilots were asked to provide comments regarding the desirability and suitability of data link gating and respond to a series of open ended questions during the post experiment debriefing. These questions included: 1) What did you like best about the system you used?, 2) What did you like least about the system you used?, 3) What potential benefits do you see for data link gating?, 4) What potential drawbacks do you see associated with data link gating?, and finally 5) Should pilots retain control over activation of all autopilot modes? If no, which modes if any could/should be activated automatically? Pilots in the no gating conditions were shown a demonstration of a data link gating system prior to responding to this worksheet in order to provide some basis for their responses.

10.7.1 *Data link desirability.* Pilots were asked to respond to the statement: "I think that gating capability should be included in future data link systems" and provide written reasons for their responses. Table 20 indicates the median and range of responses broken down by experimental condition. In general, pilots were enthusiastic about data link gating. A Kruskal-Wallis one way ANOVA showed that the differences between experimental conditions was not significant ($X^2_{KW} = 3.12$, $p = .2106$).

Table 20. Responses to the statement: "I think that gating capability should be included in future data link systems" (1= agree, 7 = disagree)

Condition	Median	Range
No Gating	2.25	1.0 – 4.0
Gating/Text	1.75	1.0 – 6.0
Gating/Graphic	3.875	1.0 – 6.0

The reasons most frequently given for the generally expressed desirability of gating are shown in Table 21. In addition to these comments, two pilots commented that while they generally favored data link gating, they felt data link gating should only send data to the MCP and not the FMC due to the limited visibility of the FMC CDU. Also, two pilots indicated that while they generally favored data link gating, they did not feel that it was appropriate for the terminal area, due to the fast pace of operations.

Table 21. Reasons for desiring gating in future systems (n = 15 respondents)

Reason for Desiring Gating	*n*
Decrease in workload	7
Time savings	3
Decreased frequency congestion	3
Decreased Errors	2

10.7.2 *Best and least liked data link features*. Table 22 depicts the relative frequency of best-liked system features broken out by gating/display condition. The responses did not vary greatly across experimental conditions (pilots in the no gating condition could not comment on the benefits of gating). The data link display received the most positive comments, the majority of which cited the memory aid benefits of permanently displaying the current ATC clearance. In addition to commenting on the specific data link systems used in this study, pilots were also asked about the general benefits of data link gating. Workload benefits accounted for the majority (19/36) of these comments.

Table 22. Best liked data link features (n = 28 respondents)

Best Liked Feature	No Gating	Gating/Text	Gating/Graphic	Total
Data Link Display				
Memory aid	7	6	3	16 (40.0%)
Error reduction	3	2	1	6 (15.0%)
Gating				
Workload reduction		6	6	12 (30.0%)
Convenience		1	2	3 (7.5%)
Miscellaneous		1	2	3 (7.5%)
Total	10	16	14	40

Table 23 indicates the most commonly cited problems with the data link systems used in this study. Most pilot complaints (64%) centered on various features of the data link display, while complaints regarding data link gating were less frequent (18%). Pilots expressed a desire for more information on the data link display which, like all currently proposed data link displays, only displayed the current data link clearance. Previous clearances could be reviewed only via a clearance log. Pilots indicated that even though the log was available, any constraints contained in previous clearances (especially altitude/airspeed constraints) that still apply should also be displayed on the data link display along with the text of the current data link clearance. In addition to these specific comments, pilots were also asked about envisioned problems with data link gating in general. Twenty five percent of these comments indicated potential problems with over-trust and complacency. Other comments included concerns over monitoring workload (6/48 comments) and loss of situation awareness (6/48 comments).

Table 23. Least liked data link system features (n = 26 respondents)

Least Liked Feature	No Gating	Gating/Text	Gating/Graphic	Total
Data Link Display				
Wanted more information on DL display	4	2	5	11 (28.2%)
Didn't like display format/color	4	4	0	8 (20.5%)
Display too far from ND/PFD	3	0	3	6 (15.4%)
Gating				
Reduces pilot involvement		2	2	4 (10.3%)
FMC gating is difficult to verify		0	2	2 (5.1%)
Makes negotiation difficult		0	1	1 (2.6%)
Miscellaneous		4	3	7 (17.9%)
Total	11	12	16	39

10.7.3 *Control over mode activation.* In this study, pilots were required to activate all modes; however, in future systems, mode selection could also be delegated to the automation. During the debriefing, pilots were asked which modes, if any, should be automatically activated. Not surprisingly, a vast majority of pilots responding to this question (82%) preferred to retain control over mode selection (see table 24). Those pilots who felt that mode selection could be automated to some extent indicated that automatic selection of either Heading Select or LNAV would be acceptable. Unlike the selection of a vertical mode which involves at least three different options, there is little flexibility in selection of a lateral mode. A clearance that calls for a given heading or routing change will always necessitate selection of Heading Select or LNAV modes respectively. Therefore, automatic mode selection of either Heading Select or LNAV may minimize failures to activate the appropriate mode.

Table 24. Pilot responses to the question "Should pilots activate all autopilot modes?"

Response	*n*	%
Pilot should activate all modes	23	82.1
Heading Select mode should be automatic	3	10.7
LNAV mode should be automatic	1	3.6
Both Heading Select and LNAV should be automatic	1	3.6
Total	28	100

11.0 Discussion

The evolution of automated systems from reactive tools to highly capable and autonomous agents has resulted in an increased need for human-machine coordination. Intentions and actions need to be communicated and agreed upon, and resources need to be allocated in a timely and efficient manner. One possible strategy for achieving this goal is the so-called management-by-consent approach where operators retain ultimate control of the overall system as machines cannot take any action unless and until explicit operator consent has been received. A major challenge for this approach is to ensure that the human is assisted in providing informed (rather than perfunctory) consent by supporting the timely detection of conflicts between machine goals and activities and the constraints imposed by tasks and the environment. This study examined the effects of type of conflict, time pressure, trust, and display design on conflict detection processes and performance. The context for this research was data link, the proposed medium for digital air-ground communication and coordination in the aviation domain. In particular, this study examined the impact of its capability for the direct transfer or "gating" of information on human-machine coordination in the automated flight deck system. The following sections will discuss in detail the findings of this study, its implications for human-machine coordination, and possible methods to improve conflict detection.

11.1 *Conflict detection performance*

Pilots have expressed a strong preference for a management-by-consent approach (compared to more highly automated approaches) due to the perceived high level of operator control under this coordination strategy (Olson and Sarter, 1998). However, the results of the current study indicate that one important prerequisite for effective control, namely reliable and timely conflict detection and resolution, is not necessarily supported by modern technology. Fewer than half of the goal conflicts (42.5% for the no gating condition and 27.5.% for the gating conditions), and none of the implementation conflicts in this study were detected prior to the consent decision (see figure 11). After the consent decision, overall conflict detection was somewhat improved, but was still relatively poor with only 75% of goal conflicts and 50% of implementation conflicts being detected (see figure 13). Note that this study focused on differences in conflict detection between different *data link* systems; conflict detection using *voice transmittal* of ATC clearances (the current communication method) was not investigated. However, earlier research by Hahn and Hansman (1992) indicates that conflict detection with existing voice communication is comparable to detection performance when using a text based data link display (a display very similar to the one used in the no gating condition in the current study).

In an attempt to understand the reasons underlying the observed conflict detection performance in this experiment, a variety of potential influencing factors were examined. Conflict type, time pressure, gating, and operator trust all had a significant impact on detection performance, while display design did not have an effect. The following sections will briefly reintroduce the model of conflict detection shown in table 7 and use it to help describe and explain the observed effects of these factors in more detail.

11.1.1 *The conflict detection model.* This study used verbal protocol and observational data to better understand the processes and stages involved in conflict detection (see table 7). Our analysis shows that, *prior to the consent decision,* conflicts were detected primarily through knowledge-based evaluation. The operator read the data link clearance and compared the expected results of accepting and loading the clearance with knowledge of current aircraft state and other goals, activities, and constraints associated with desired aircraft performance.

After the consent decision, conflicts were detected via three different processes. During post-consent confirmation (following clearance acceptance and loading), pilots typically reread the data link clearance, checked the MCP and/or FMC CDU to ensure that the expected targets had been loaded correctly, and checked specific cockpit displays to confirm that the aircraft was *initially* responding in the desired manner. During subsequent monitoring, conflict detection involved both expectation-driven and data-driven processes. Expectation-driven conflict detection was typically preceded by rereading the data link clearance and involved scanning the specific cockpit instruments to confirm that system behavior matched pilot expectations. Data-driven monitoring was typified by a general scan of cockpit instruments not specifically related to the previous clearance instructions. Some unexpected or undesired indication would catch the pilot's attention and lead to the detection of the problem. If conflicts were not detected by any of these processes, forcing functions (the requirements of other tasks) could result in conflict detection. The importance of this stage will be discussed in more detail in subsequent sections.

An examination of the various stages in this model shows that conflict detection depends heavily on pilot expectations regarding data entry and system response. This is in agreement with previous automation research (Sarter and Woods, 1995, 1997) which has indicated that adequate operator knowledge of and expectations regarding system behavior are vital to the effective use and monitoring of automated systems. The observed effects of conflict type, time pressure, and gating may be attributed to pilots' difficulties with generating adequate expectations to guide their decision-making behavior. Specific examples of these difficulties will be discussed during the analyses presented in the following sections.

11.1.2 *The effects of conflict type*. In general, implementation conflicts were less likely to be detected than goal conflicts. Within these general categories, inappropriate conflicts were detected less often than impossible conflicts, and those in which automation did more than expected were less likely to be detected than conflicts in which the automation did less than anticipated.

Goal vs implementation conflicts. Previous data link research (Hahn and Hansman, 1992, Logsdon, 1996) has examined only the detection of goal conflicts, whereas this study examined both goal and implementation conflicts. As predicted, this study found that implementation conflicts were more difficult to detect than goal conflicts both before the consent decision (32.5% goal vs 0% implementation conflicts detected) as well as overall (75% goal vs 49% implementation conflicts detected) (see figures 11 and 13).

These conflict detection differences are largely due to the extra cognitive effort required to evaluate the implementation of the clearance by two machine agents. In order to assess the appropriateness of clearance implementation, the pilot had to consider: a) the performance targets generated by the data link system, b) the intended destination of those targets, and c) how they would be interpreted by the MCP and FMC. This task was complicated by data link displays that indicated only the value, but not the destination, of performance targets. Furthermore, the pilot had to consider how the implementation method chosen by the data link system would interact with the aircraft mode to be selected by the pilot. In many cases, the same data link clearance could be implemented using one of several different data entry methods, each of which required the selection of a different mode. For example, a given altitude restriction could be entered in either the MCP or the FMC CDU. If the selected mode did not match the location of data entry, the aircraft would not comply with the desired altitude restriction. As a result of the cognitive effort required to evaluate these considerations, and since automated systems do not necessarily follow the same procedures as their human operators (Sarter and Woods, 1994, 1997), implementation conflicts were often difficult to detect.

Inappropriate vs impossible goal conflicts. – Two types of goal conflicts were examined in this study. Inappropriate conflicts resulted when the contents of the clearance conflicted with other pilot goals, procedures, or regulations (e.g., a heading that would place the aircraft too close to the runway on final approach). In the case of impossible conflicts, the text of the data link clearance conflicted with the current situation (e.g., a clearance to descend to an altitude above the current altitude). Previous research has not addressed the relative detection differences between these two goal conflict types. While Logsdon (1996) presented both types of conflicts, she did not report the difference in detection performance.

Contrary to our predictions, inappropriate conflicts were significantly more difficult to detect than impossible conflicts prior to the consent decision (15% of inappropriate vs 50% of impossible conflicts). This detection difference may be explained by the fact that the detection of inappropriate and impossible conflicts required different comparison procedures. In most cases, the detection of inappropriate conflicts involved a comparison of the clearance contents with applicable flight regulations and procedures residing in long-term memory. In contrast, detection of impossible conflicts required a comparison with the current state of the aircraft which resided either in working memory or was presented on cockpit displays. In other words, detection of inappropriate conflicts required the use of knowledge in the head, while detection of impossible conflicts required use of knowledge in the world or in working memory (Norman, 1988).

Previous work (Norman, 1988) indicates that the activation of "knowledge in the head" can be relatively effortful and slow. The analysis of conflict detection activities observed in this study seems to support this finding. Of the 48 impossible conflicts detected, 98% were detected during the pre-consent evaluation (69%) and post-consent confirmation (29%) stages. In contrast, of the 42 inappropriate conflicts detected, 50% were not detected until the last two conflict detection stages - subsequent monitoring (31%) and detection via forcing function (19%). These results suggest that pilots were able to more quickly deploy the knowledge in the world required to detect impossible conflicts.

Implementation does more vs implementation does less. This study also examined detection differences between the two types of implementation conflicts. Conflicts in which automation did more than expected were far less likely to be detected (26%) than conflicts in which automation did less than expected (73%) (see figure 13). This result confirms the findings of earlier research (Sarter and Woods, 1997) and can be explained by the expectation-driven nature of both the post-consent confirmation processes and subsequent monitoring processes.

At the post-consent confirmation stage, pilots referred to specific cockpit displays to ensure that the data link system had transferred the correct data to the FMC CDU and MCP and that the system behavior was initially proceeding in accordance with expectations. Subsequent monitoring was also largely expectation-driven. Pilots were frequently observed to reread the text of the data link clearance and then check specific cockpit displays to confirm that system behavior matched the instructions in the ATC clearance. Consequently, conflicts in which automated systems did less than expected were easier to detect since pilots would quickly notice the discrepancy between expected and observed system targets and behavior.

In contrast, when automated systems did more than expected, all of the pilot's expectations were satisfied. In order to notice the additional undesired system activities, attentional guidance from the system would have been necessary, but was often hampered by the low observability of many automated systems. For example, when the data link system automatically loaded a commanded change to the landing runway but also deleted a portion of the vertical profile, the latter change was not immediately visible to the pilot due to the limited display space available on the FMC CDU (the so called keyhole property – Woods, Johannesen, Cook, and Sarter, 1994). Instead, the pilot had to select a specific page of information to detect the changes to the vertical profile – an unlikely action given the absence of corresponding expectations.

11.1.2 *The effects of* time *pressure*. In this study, time pressure was manipulated via the urgency of data link clearances. This time pressure manipulation affected only the processes involved in pre-consent evaluation. The effects of time pressure could be observed only for goal conflicts since none of the implementation conflicts were detected prior to the consent decision. As expected, fewer urgent goal conflicts were detected prior to giving consent (23% of urgent vs 41.7% of routine goal conflicts) (see figure 11).

Previous research (Edland and Svenson, 1993, Orasanu and Connolly, 1993) indicates that high levels of time pressure may result in the use of simplified strategies, decreased depth of processing, and a decrease in the amount of information gathered. The observational data from this study show that our time pressure manipulations replicated the latter two effects. While processing urgent clearances, subjects were observed to sometimes read only a portion of (as opposed to the entire) data link clearance. Pilots were also quicker to accept urgent clearances and, especially in the two gating conditions, were often unable to combine the separately understood pieces of the clearance with knowledge of aircraft state to enable conflict detection.

Additional analyses revealed an interaction between time pressure and gating for the detection of impossible conflicts. High time pressure had no effect on the detection of impossible conflicts in the no gating condition; however, subjects in the two gating conditions detected far fewer urgent clearances (20% for urgent, 60% for routine). They were more likely to detect an impossible conflict as a result of observed discrepant behavior (e.g., an observed climb vs. an expected descent) during the post-consent confirmation stage. Since detection of impossible conflicts during the pre-consent evaluation stage requires a comparison between the instructions in the clearance and knowledge of current aircraft state, this finding suggests that pilots in the gating conditions were less aware, or held a less active representation in working memory, of this information. As a result, they were often unable to deploy this knowledge under time pressure. This effect may be one manifestation of reduced operator involvement and the "out of the loop" problems (e.g. Wickens, 1992) associated with higher levels of automation. This finding also indicates that when help from automated systems is needed most (under time pressure, for example), the least amount of help is provided – another example of "clumsy automation" (Wiener, 1989).

The delayed detection of urgent impossible clearances in the two gating conditions is a matter of practical concern. When pilots in the two gating conditions finally detected these conflicts, the aircraft had already departed from the previously assigned altitude. Such altitude deviations can seriously compromise safety. In other domains, such as process control, the effects may be even more problematic as some system actions may be difficult or impossible to reverse once initiated.

11.1.3 *The effects of gating.* The results of this study replicate and also help to explain the conflicting results of previous data link studies. As discussed in the previous section, this study found that, under time pressure, pilots in the no gating condition were more likely to detect impossible conflicts than pilots in the two gating groups (see figure 12). This result partially replicates the findings of Logsdon (1996) who showed that subjects using a gating system were less likely to detect a mix of inappropriate and impossible conflicts before the consent decision.

This study also found a non significant trend towards an overall detection advantage for inappropriate conflicts in the gating condition (see figure 14). This finding replicates the magnitude of the detection advantage reported by Hahn and Hansman (1992) who suggested that eliminating the tasks associated with manual data entry freed the pilot to devote more resources to evaluating the clearance at a strategic level, thereby increasing situation awareness. The results of this study support a slightly different interpretation. The detection of inappropriate conflicts prior to the consent decision was identical for both no gating and gating conditions. Observed detection differences occurred after the consent decision and were most related to differences in monitoring effectiveness. An examination of the data indicates that 25% of all inappropriate conflicts were detected during monitoring in the two gating conditions while, in the no gating condition, only 15% of those conflicts were detected during monitoring.

This pattern of results does not fit the explanation provided by Hahn and Hansman (1992) who suggested that since gating frees the pilot from data entry tasks, it may allow the pilot to develop a better awareness of the strategic implications of a given data link clearance. The implications of this argument are that this greater strategic awareness should also lead to a detection advantage prior to the consent decision. However, the present study found no differences in the detection of inappropriate conflicts prior to the consent decision. An alternative explanation for these results derives from Gai and Curry's (1976) accumulation model in which monitoring is seen as the accumulation of evidence over time. Since pilots in the gating conditions were able to load and activate ATC clearances significantly faster than pilots in the no gating conditions, monitoring may have been more effective for pilots in the gating condition simply because they were able to spend more time monitoring.

11.1.4 *The effects of trust.* In this study, pilots provided ratings of trust in ATC, data link systems, and the FMC immediately following each scenario. These trust measures were generally positive, highly related, and showed a significant relationship with conflict detection prior to consent. These findings provide important insights into the nature of pilot trust in general, as well as the relationship between trust and conflict detection.

Pilot trust in ATC and automated systems was moderately high and relatively insensitive to problems with clearance goals or implementation (see figure 18). Median trust ratings indicated a generally high level of trust (between 5 and 6 on a seven point scale), that either did not change (trust in ATC) or increased slightly (trust in data link systems and the FMC) over the course of the study. This finding confirms some earlier research (Gempler and Wickens, 1998,; Conejo and Wickens, 1997) showing that implicit measures of pilot trust did not significantly decrease with failures of an automated cockpit cueing device. However, this result conflicts with research by Lee and Moray (1992) which showed significant decreases in subjective trust measures following failures in an automated process control task. There are two potential explanations for the findings of the current study. First, the relatively stable and generally high levels of trust expressed in this study may be due to extensive and generally positive previous pilot experiences with ATC and automated systems (in contrast to the relatively small amount of experience accrued by subjects in Lee and Moray's (1992) process control study). Since trust is related to perceived competence (Muir, 1988), the relatively small number of actual problems in this study (one out of five clearances contained a conflict, and an even smaller number were actually detected) may not have been sufficient to affect perceived competence (and therefore trust) when considered against the backdrop of thousands of previous experiences. Also, the fact that subjects' trust was below the maximum rating ('7) may indicate that it was fairly well calibrated to the actual error rate (Gempler and Wickens, 1998). Second, the generally high trust in data link systems (with which pilots had no previous experiences) may be explained by a transfer or generalization of high trust from the FMC and ATC to the data link system. Additionally, pilots may have blamed the FMC or ATC for problems that, in fact, were due to implementation methods chosen by the data link system.

As mentioned above, the three trust measures used in this study were highly correlated (see tables 13 and 14). This was especially true of trust in the two automated systems – data link and the FMC. This interrelation between trust measures suggests that operators in widely distributed multi-agent systems such as data link may not be able to apportion trust and assign blame in response to observed undesired system behavior, especially between automated systems. In other words, the relationship between trust measures may, to some extent, indicate a potential problem with responsibility awareness (Coury and Semmel, 1996). Observational data from this study give some indication of the type of potential problems that may be caused by inadequate responsibility awareness. For example, in one scenario, poor implementation by the data link system and FMC resulted in the undesired deletion of an altitude constraint within the FMC. In one case, a pilot detected this undesired action, but was unsure whether this problem was caused by the unintended actions of the automated systems, or was the result of an intended (but poorly communicated) instruction from ATC. As a result, instead of immediately reprogramming the FMC CDU to correct the problem, the pilot first contacted ATC and asked what the controller really wanted. In this case, a lack of responsibility awareness resulted in a delayed correction of undesired system behavior.

Finally, trust in automated systems and ATC was related to conflict detection prior to the consent decision (see table 15). Pilots indicating low trust in ATC were more likely to detect conflicts prior to the consent decision. Since only goal conflicts (those related to problems with the clearance itself) were detected prior to the consent decision, this finding is in line with our prediction that trust in ATC should be associated with detection of goal conflicts. Due to the relatively small number of conflicts detected prior to the consent decision, it is difficult to determine the specific effects of low trust in ATC on underlying detection processes. Presumably, low trust in ATC led to a more thorough evaluation of the data link clearance which in turn led to a greater likelihood of detecting goal conflicts prior to the consent decision. This finding also confirms previous research which has shown a relationship between operator trust and the use and monitoring of automated systems (Moray and Lee, 1996; Parasuraman, Molloy, and Singh, 1993; Lee and Moray, 1992).

Trust in automated systems was also related to detection of goal conflicts prior to consent, but in the opposite direction. Pilots who indicated the highest levels of trust in automated systems were more likely to detect conflicts prior to the consent decision. Although the data do not support any one particular explanation, it may be that subjects with high trust in automated systems spent relatively less effort evaluating possible implementation problems and therefore could spend more effort evaluating the acceptability of the ATC instructions.

There was no relationship between trust and overall conflict detection. This lack of significant relationship may be explained by the variety of stages that contribute to overall conflict detection. While detection prior to the consent decision is the result of only the pre-consent evaluation stage, overall conflict detection also includes post-consent confirmation stage, subsequent monitoring, and detection via forcing function. While trust in ATC and automated systems may have specific effects at any one stage, when performance across all stages is considered, these effects may be offset resulting in no significant overall relationship.

11.1.5 *The effects of experience.* Pilots with the greatest overall flight experience (> 10,000 hours) as well as moderate levels of experience in the B757 (600 – 1500 hours) were least likely to detect conflicts (both overall as well as prior to the consent decision) (see table 16). Although neither the data nor previous research support any one particular explanation, these results may be due to difficulties generating expectations regarding machine actions required for conflict detection. Pilots with moderate levels of experience in the B757 may have forgotten some aspects of system operation learned in initial training and not yet acquired a sufficient set of personal experiences to replace that knowledge. Overall experience may also affect pilot expectations. Since glass cockpit aircraft are a fairly recent phenomena (dating from the early 1980s), highly experienced pilots have acquired the majority of their flight experience on less automated aircraft. Previous research (Wiener, 1989) indicates that many older captains are less comfortable with cockpit automation and often defer to more junior crew members in the operation of these systems. As a result, pilots with the most overall flight experience may have difficulties generating their own expectations of automated system actions and thus be less able to detect conflicts.

11.1.6 *The effects of Display design.* In contrast to our predictions as well as previous data link research (Hahn and Hansman, 1992), the use of graphic displays did not lead to superior conflict detection performance in this study. This may be explained, in part, by the Proximity Compatibly Principle (Wickens and Carswell, 1995) which states that display and task proximity effects are moderated by factors such as information access cost, clutter and confusion, and the presence of emergent features. In this study, the information access cost associated with the distance between the data link text display and the displays of data link information on the ND/VSD in the graphic condition may have contributed to the observed lack of conflict detection benefits. In the debriefing, many pilots complained about the distance between the data link clearance displayed on the right monitor and the display of data link information on the ND and VSD (left monitor) which may explain the fact that, during the pre-consent evaluation stage, pilots referred to information on the both displays only infrequently. Note that the relative dispersion of these displays in this study was not the result of a poor design choice on our part, but was designed to conform to the location of available display space in actual aircraft and to currently proposed data link display positions.

Information access cost may also explain the failure of this study to replicate the conflict detection advantage for graphic displays found by Hahn and Hansman (1992). In the Hahn and Hansman study, information access cost associated with the graphic display was relatively low because in their graphic display conditions, there was no display of the text of the data link clearance. Instead, the data link clearance was depicted graphically on the ND and PFD (only) by the use of lines and pointers. This presentation reduced information access cost since the data link clearance information was superimposed over the existing flight instruments. While this type of graphic display minimizes information access cost, it involves its own problems. Not all data link clearance information can be depicted graphically. In particular, semantic information such as "until advised" or "cleared for the approach" is difficult to present in a graphic manner.

11.2 *Implications of Our Findings for Human-Machine Coordination*

Coordination theory (Malone and Crowston, 1990) describes coordination as the management of dependencies between agents. This process requires that the human operator detects dependencies and conflicts in the first place. The results described in previous sections indicate that pilots are often unable to meet this requirement. Their poor detection performance both prior to and after the consent decision may be understood by examining human-machine roles and coordination demands in supervisory control systems.

At a general level, the gating system employed in this study represents the automated transfer and translation of external goals into a set of system commands to different elements of a supervisory control system. In order to extend the findings of this study beyond the specific data link system employed as well as to better understand the underlying causes of poor conflict detection performance, the following sections will describe the general effects of gating on operator roles and coordination opportunities.

11.2.1 *The new role of the operator.* Figure 19 indicates the relationships between human and machine agents in a glass cockpit aircraft based on Sheridan's (1997) general model of supervisory control. The current (no gating) system is depicted in the left hand panel, while the effects of gating are indicated in the right hand panel.

In the current system, the human operator (pilot) responds to a requested change in system goals (e.g., a pending ATC clearance) by instructing the automation via a set of performance targets provided to the Human Interactive Computers (HICs) (in this case, the FMC and MCP). The HICs translate these higher level goals into a set of commands sent to lower level Task Interactive Computers (TICs) (the autopilot and autothrottles) which communicate with the mechanical actuators that produce actual system performance. The operator then monitors resulting system performance and intervenes when observed performance deviates from desired performance. In this type of system the operator coordinates human and machine goals and actions by 1) instructing the HICs, 2) selecting the desired operating mode, and 3) intervening in the event of undesired performance.

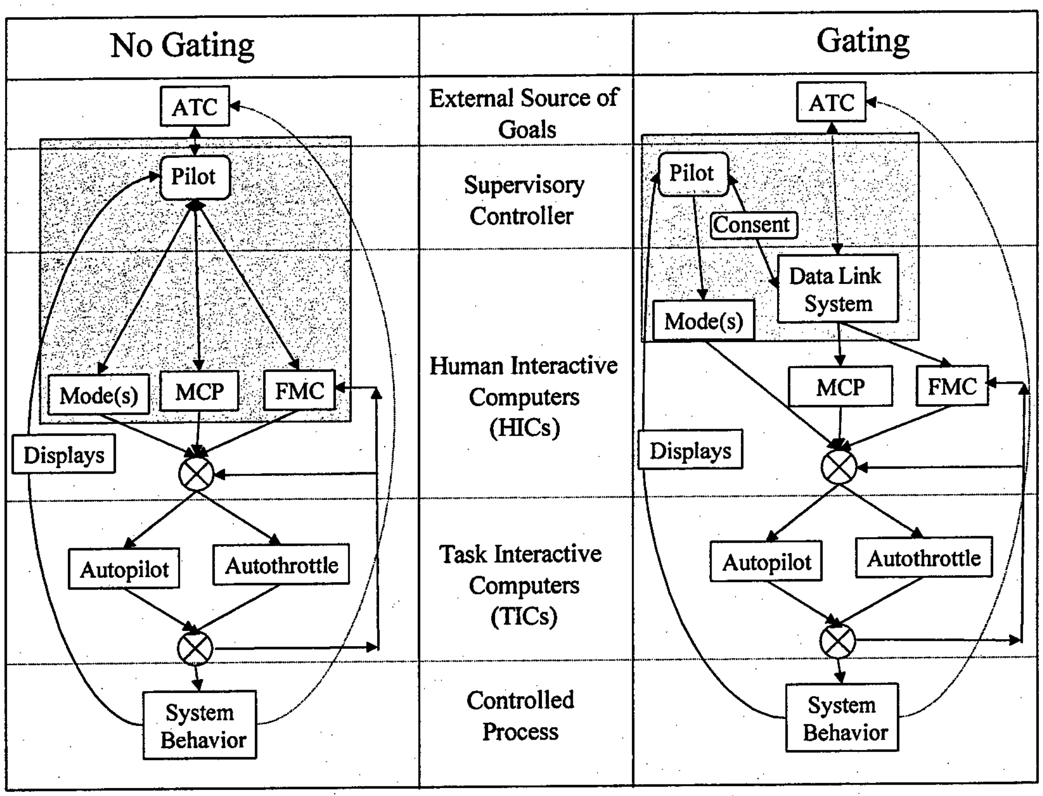


Figure 19. The glass cockpit as a supervisory control system (adapted from Sheridan, 1997).

Automated system features such as data link gating are intended to increase system efficiency by automating the planning and instruction processes. The shaded box in the right panel of figure 19 indicates the changes to the operator's roles in such a system. In this case, the request for a change in system goals is relayed to the pilot via a digital communication system (data link). After receiving operator consent, the system loads the targets associated with these commands. Note, however, that the human must still select the desired operating mode. In this type of a system, the operator coordinates human and machine goals and actions by 1) providing consent to the goals and implementation methods sent to the HICs, 2) by selecting the desired operating mode, and 3) by intervening in the event of undesired performance. As indicated by a comparison of the left and right panels of figure 19, gating adds one more level of automation above the HICs and further reduces operator involvement in system control. This system also splits the task of responding to the data link clearance between the automation, which provides the performance targets, and the pilot, who selects the operating mode. In this study, task splitting led to problems when pilots forgot to engage the operating mode, or selected an operating mode that was incompatible with the location of performance targets loaded by the automated systems.

11.2.2 *Coordination and shared cognition.* The model described in figure 19 points out the distributed nature of coordination in supervisory control systems. Instead of coordinating with one monolithic automated system, the human operator is responsible for coordinating with and among a number of related, yet separate machine entities (the HICs and TICs). As previously noted, breakdowns in human-machine coordination often occur because these various human and machine agents possess imperfect knowledge of each other's intentions and actions (Suchman, 1987; Roth, Bennett, and Woods, 1987). Additionally, breakdowns in machine-machine coordination are often observed because the machine components of supervisory control systems possess imperfect knowledge regarding the abilities and intentions of other machine components (Moray, 1986). In essence, the structure of supervisory control systems requires machine systems at successive levels to interpret the instructions passed along by higher level components. In this situation, as in human communication, when components do not possess shared knowledge of abilities and intentions, breakdowns in coordination are likely (Clark and Brennan, 1992). As a result, coordination in supervisory control systems requires the operator to not only coordinate machine actions with operator goals and activities, but to also ensure coordination *between* machine agents.

The introduction of gating systems increases the coordination demands on the operator because it forces the operator to understand the interpretation of system goals by yet another agent - the gating system. Additionally, since gating in currently proposed data link systems does not directly control system behavior, but only passes commands to the FMC and MCP, the operator must also understand how these components will interpret the targets sent by the data link system. In this study, implementation conflicts were the result of a lack of shared understanding between the FMC and MCP rather than the result of an improper translation of the data link clearance by the gating system. For example, in one implementation conflict the data link system translated a clearance to change landing runways into a command to change the selected runway on the appropriate page of the FMC. Due to a lack of shared knowledge, the FMC was unable to correctly interpret this command. As the result of data propagation within the FMC, this runway change command also resulted in a change to the altitudes used to compute the descent profile. Since the FMC did not know (and could not ask) whether the pilot/data link system intended to retain the previous descent profile, the system assumed that a new descent profile would be entered and deleted the previous profile resulting in an implementation conflict.

11.3 *New opportunities for error*

The literature on human machine interaction suggests that, while automated systems deliver many promised safety and efficiency benefits, they often also create new human-machine coordination problems and new opportunities for error (Woods, 1996). Data link gating provides yet another example of this phenomenon. The analysis of operator roles in the previous section indicates that gating relieves the pilot of programming and instructing tasks, but adds to coordination demands by forcing the operator to understand the interpretation of the ATC clearance by the gating system. The results of this study indicate that this change in roles reduced operator workload, decreased response time, avoided data entry errors, and improved conflict detection in some cases. However, gating also led to reduced conflict detection in some cases, imposed new attentional and knowledge demands on the operator, and provided new pathways for error in the form of undetected implementation conflicts and mismatches between data entry and mode selection.

11.3.1 *Benefits of data link gating.* Data link gating has been proposed as a method of decreasing pilot workload and data entry errors (Knox and Scanlon, 1990). As indicated by superior secondary task performance as well as subjective comments, in this study gating indeed decreased the workload associated with instructing the FMC and MCP. It also reduced data entry errors, allowed faster response to data link clearances, and led to superior detection of inappropriate conflicts. Each of these benefits will be briefly discussed in the following sections.

Reduced data entry errors. Gating should eliminate most data entry errors since the operator will only be required to program the FMC and MCP in order to correct undesired performance. Performance and subjective data support this analysis. Pilots in the two gating conditions made no FMC data entry errors, and never failed to "accept" a data link clearance. In contrast, the majority of errors made by pilots in the no gating condition were failures to "accept" the data link clearance (15/36 errors) and data entry errors to the FMC and MCP (16/36 errors). It should be noted that these error tallies do not include the numerous errors made by pilots in the urgent "implementation does less" scenario in which only 40% of the pilots in the no gating condition successfully loaded the conflicting clearance. Two pilots in the no gating condition even entered commands that would have resulted in the aircraft executing a 180^0 course reversal (had the experimenter not intervened at the last minute). Assuming that the data link system could successfully load difficult clearances such as this one, many of these potentially serious errors could be avoided. In addition to the performance data, subjective data also indicate that behind workload reduction, error reduction was the second most frequently cited benefit of data link gating.

Response time benefits. This study replicated previous research (Waller, 1992, Knox and Scanlon, 1991, Logsdon, 1996) which indicated that gating allowed for a quicker response to data link gating. On average, gating allowed pilots to accept and load data link clearances almost four times faster than pilots in the no gating conditions (accept – 7.48 vs 28.45 sec, load – 9.01 vs 34.18 sec). Additionally, gating allowed pilots to activate horizontal and vertical modes almost twice as fast as in the no gating conditions (with the exception of the Heading Select mode). Pilots in the no gating condition were able to activate the Heading Select mode faster since they processed data link clearances one element (i.e. heading, altitude, or airspeed) at a time. Since heading instructions were the first clearance element, they were processed first in the no gating condition.

Conflict detection benefits. As discussed previously, pilots in the two gating conditions were able to detect more inappropriate conflicts, although this advantage was only marginally significant. It appears that the time savings and workload reduction benefits of gating allowed the pilots to detect conflicts more effectively during monitoring processes either due to an increased awareness of the strategic implications of the data link clearance, or due to the relatively longer period of time spent monitoring.

11.3.2 *Costs of data link gating.* While gating provided promised benefits, it also changed the nature of human-machine interaction which led to new attentional and knowledge demands and created new opportunities for error. Indications of each of these problems will be briefly discussed in the following sections.

New attentional and knowledge demands. As discussed previously, under a management-by-consent approach with direct machine-machine communication, the human operator provides to consent to not only system goals, but also to the translation of those goals into a set of system commands. In order to provide informed consent, the operator must understand the acceptability of the goals embodied in the instructions, the translation of those instructions by the mediating system, and the way in which the lower level automated systems will interpret and execute these goals. Additionally, since the operator did not translate or enter the commands into the system, he/she must be able to guide attention to the appropriate display locations to confirm data entry following the consent decision.

Data link gating is an example of a system that imposes these new knowledge and attentional demands on the operator. However, it does not reduce substantially the knowledge and attentional demands imposed by manual data entry. In spite of automated data entry, the operator must still understand how the FMC and MCP will interpret commands passed on by the data link system in order to anticipate and evaluate the acceptability of the chosen implementation method and select the appropriate aircraft mode(s). For example, the data link system could send airspeed commands to either the MCP or the FMC CDU. In order to comply with an assigned speed restriction, the pilot would have to know or identify where targets were sent and then choose the vertical mode appropriate for the implementation method selected by the data link system. In addition, knowledge of manual data entry methods is also necessary to intervene and resolve problems arising from improper data entry or a changing situation.

Both the performance and the subjective data from this study indicate the effects of these new attentional and knowledge demands. Pilots were unable to deploy the knowledge and attentional resources required to detect <u>any</u> implementation conflicts prior to the consent decision even though the automated systems in this simulation acted just like the systems on their aircraft. Additionally, the relative inability to detect conflicts in which automated systems did more than expected reflects the need for, but lack of, support for data driven monitoring. Pilot's subjective comments also attest to the effects of these new attentional and knowledge demands. They cited new monitoring demands and an inability to monitor system actions as potential problems with data link gating.

<u>New opportunities for error.</u> The introduction of new highly automated systems often results in the creation of new error types and pathways (Woods, 1996). These new

opportunities for error arise from the changes in operator roles and coordination demands discussed in previous sections. The pattern of errors and undesired system events observed in this study illustrates this process. For example, gating automates the majority of data entry, and therefore the reduction in data entry errors is not entirely surprising. Note, however, that gating did not eliminate these errors altogether, it merely changed their nature and underlying causes. In the no gating condition, the pilot was the source of the data entry errors. In the case of the two gating conditions, undesired events occurred when targets were entered in an unexpected way or location by the data link system or were misinterpreted by the FMC. Thus, data entry errors associated with the gating system represent coordination failures between humans and machines rather than the slips or mistakes usually associated with manual data entry.

This study also provides further support for the claim that increasing levels of system autonomy will lead to a shift from errors of commission to errors of omission due to operator difficulties predicting and tracking automated system behavior (Sarter, and Woods, submitted). As indicated by the results of this study, poorly formed operator expectations often lead to a failure to intervene in undesired system action –an error of omission – rather than the errors of commission associated with the control of less automated systems.

Gating also created new opportunities for error during mode selection which remained the operator's task. Aircraft behavior is dependent on both the mode selected by the pilot as well as the data entered by the gating system. As a result, if the pilot failed to track or consider the data entry method chosen by the data link system, the aircraft would not behave as expected by the pilot. In the two gating conditions, mismatches between data entry and mode selection accounted for the largest percentage of deviations from ATC clearances (36%).

11.4 *Possible Ways to Improve Conflict Detection*

One contribution of this study is that it provides a model of pilot conflict detection activities and processes that can be used to not only better understand the reasons behind observed performance effects, but also to suggest specific countermeasures. Similar to Sellen's (1994) analysis of human error detection mechanisms and Reason's (1990) conception of defenses in depth, the model of conflict detection developed in this study can be conceived of as a series of safety nets that allow for conflict detection to occur. Undesired system events represent those conflicts that escape detection at all stages.

The results of this and earlier studies indicate that conflict detection relies primarily on expectation-driven comparison and evaluation processes. Conflict detection failures are largely the result of inadequate expectations regarding expected system behavior or low observability. In order to improve conflict detection performance, possible solutions include external attentional guidance during post-confirmation and monitoring, a reduction of coordination demands, and the consideration of the importance of other agents in the conflict detection process. The following sections will describe specific recommendations within these general categories.

11.4.1 *Support for external attentional guidance.* The results of this study indicate that conflict detection relies heavily upon the presence of adequate operator expectations required to predict the results of pending machine action prior to consent, to confirm data entry following consent, and to drive the allocation of attention during subsequent monitoring processes. In the absence of adequate expectations, external attentional guidance may substantially improve the operator's ability to detect conflicts at each of these stages (Sarter, Woods, and Billings, 1997). Data from this study suggest several specific methods of guiding attention.

Visualization of goal translation. One of the difficulties posed by gating is that it requires the operator to understand both the translation of external goals into a set of system commands as well as the interpretation of those commands by other automated systems. In order to reduce the ambiguity associated with these processes, the data link display should indicate the performance targets sent by the data link system, the intended destination of those targets within the FMC and MCP, and the predicted aircraft performance that would result from loading the data link clearance. This type of display would: 1) reduce the cognitive effort required to evaluate the acceptability of loading a data link clearance prior to consent, and 2) assist in directing operator attention to the location of relevant data during post-consent confirmation processes.

Long-term availability of gated data location. The results of this study indicate that, immediately after loading a data link clearance, pilots engaged in expectation-driven monitoring processes to ensure that expected data were loaded and the initial system response corresponded to expectations. In the absence of operator expectations, conflicts were rarely detected at this stage. As a result, detection of conflicts in which automated systems did more than expected was extremely poor. This finding indicates the importance of indicating the locations (i.e. MCP or location within the FMC CDU) of gated data through the duration of the post-consent confirmation stage.

Highlight locations of gated data. The basic attention literature suggests that attention allocation is influenced by both operator expectations and visual display properties (e.g. Egeth and Yantis, 1997, Jonides and Yantis, 1988). In addition to the above mentioned solutions which affect information content, changing the display properties of existing information may also support conflict detection by guiding operator attention to relevant display locations. Highlighting (through color coding, brightness, etc.) newly entered targets on the MCP and FMC would help draw operator attention to these values during the post-consent confirmation stage. Since the limited display space on the FMC CDU makes it impossible to readily indicate the location of all changes to command targets, an index page listing and providing links to or stepping the operator through all pages containing changes would assist the operator.

Integrated data link display. The Proximity Compatibility Principle (Wickens and Carswell, 1995) predicts that information access cost will moderate the influence of other display effects. Pilots in this study indicated that the distance between the data link display and other cockpit displays decreased the potential usefulness of data link displays. Several pilots commented that they would prefer the text of the data link clearance displayed across the bottom of the Navigation Display (ND) to facilitate evaluation and monitoring processes. The displays in this study were designed to conform to the limited display space available on existing aircraft. Our findings indicate that currently proposed display locations may not be adequate. Future data link systems should consider the practicality of incorporating data link displays into or near the Navigation Display (ND).

Display of all valid constraints. In this study, the visual display of the contents of the current data link clearance served as a useful externalization and reminder of expected/requested system behavior. Pilots were frequently observed to reread the data link clearance and then check specific cockpit indications to ensure that aircraft performance matched the current clearance. Subjective comments also indicated that the data link clearance was considered a valuable memory aid.

However, as indicated by pilot comments and observed performance, there were occasions in which it would have been useful to display some of the contents of *previous* data link clearances as well. Previous clearances were available only through a clearance log feature. This system feature mirrors the operation of most currently proposed data link systems. When the data link display did not indicate still-valid previous constraints, pilots tended not to monitor for, and therefore did not detect, violations of these constraints. For example, if a pilot was previously given an altitude restriction, followed by a subsequent clearance containing a only simple radio frequency change, the data link display only indicated the frequency change instructions. As a result, the pilot often no longer monitored for compliance with the previously assigned altitude restriction. In order to better support expectation-driven monitoring, the data link display should depict all restrictions that apply to the current situation, not just the contents of the most recent clearance.

11.4.2 *Increased coordination role for machines.* In most current highly automated systems, the operator is primarily responsible for coordinating human and machine activities and goals. In order to reduce the coordination demands on the human operator, it may be possible for machine systems to assume a greater role in human-machine coordination. Since coordination requires a knowledge of the actions and intentions of other agents, the first step in this direction is the development of systems that can infer operator intent and track operator actions. Examples of work in this area include the automated tracking of task and goal described by Jones, et al. (1991) as well as the Agenda

Manager (Cha and Funk, 1997). These systems have been shown to provide valuable assistance in prioritizing and initiating tasks. This technology could be adapted to assist in the detection of goal and implementation conflicts in supervisory control systems.

11.4.3 *Consider detection as a team effort.* In this project, the individual pilot was solely responsible for detecting and resolving all conflicts (whether planned or self-induced), no other crew members were present, and the ATC controller (the experimenter) intervened only to correct errors and deviations when necessary to remain within the confines of the experimental scenarios. In the actual flight domain, the other crew member(s) and ATC also assist in detecting and correcting errors and conflicts. It is likely that the addition other flight crew members and ATC would result in improved conflict detection and a lower frequency of undesired system events.

Previous research indicates that ATC may make the greatest contribution to error and conflict detection. A study by Mosier and Skitka (1998) indicated that relatively few errors were detected by a second flight crewmember. Also, Alexander (1998) found that ATC was responsible for detecting the majority of errors involved in incidents reported to the Aviation Safety and Reporting System (ASRS), a nationwide database of aviation incidents. These findings point to the importance of increasing the communication and coordination between human crew members in highly automated aircraft and suggest that the error detection functions of ATC must be considered in the design of ATC roles and procedures in future air traffic systems. In particular, some proposed air traffic systems may decrease the human controller's knowledge of pilot intentions either through direct communication between automated ATC and cockpit systems (Palmer, Prevot, and Crane, 1997), or by shifting responsibility for traffic separation from ground based ATC to the flight deck in the case of a proposal known as Free Flight (RTCA, 1995). A consideration of ATC's role in error detection prior to making such changes may allow designers to counteract some of the new pathways for error that these proposed changes to the future air traffic system may create.

12.0 Conclusion

As systems become more powerful, independent, and interconnected, there is an increasing need for effective human-machine coordination. One strategy for achieving this goal – management-by-consent – is considered by many to be a particularly appealing approach to the problem. The assumption is that, under management-by-consent, the human has a very high level of control over machine actions which seems critical as long as the human still bears the ultimate responsibility for system performance and the automation does not share commitments. However, the results of this study indicate that management-by-consent does not guarantee effective control. The complexity, coupling, and low observability of many automated systems can make it impossible for operators to give informed consent to machine goals and actions. Instead, they sometimes explicitly but unknowingly agree to undesirable or even unsafe system activities. Part of the problem is that, with advanced technology, the human operator needs to evaluate not only the appropriateness of one single goal for an individual system but rather has to consider the interactions and distribution of data between and within systems as well as the machine strategies for executing proposed plans of action. These strategies do not necessarily match the operator's expectations nor are they communicated effectively by the system which can lead to breakdowns in coordination and a lack of responsibility awareness. The challenge is not just for the automation to provide additional information (i.e., increase data availability) but to reduce the cognitive effort required to locate and interpret that information (i.e., improve system observability). The system needs to highlight changes in its goals and activities to support data-driven monitoring without adding confusion and clutter which could detract from other tasks. Given our findings, this seems particularly important in cases where the operator needs to evaluate machine proposals under time pressure.

In summary, observed breakdowns in human-machine reflect the inability of current automated systems to play an active role in the collaborative process of negotiating and managing goals and resources. These systems possess high levels of autonomy and sometimes act based on (false) assumptions of operator intent. However, they fail to inform their human supervisors about (difficulties with) their interpretations of commands, about progress towards shared goals, and about problems with performing an assigned task. This knowledge is critical for the operator to be able to realize the need for intervention or support of the automation and for learning about the idiosyncrasies of the machine agent in the interest of making coordination less effortful and forming a true human-machine team.

References

Aeronautica Civil of the Republic of Colombia (1996). *Aircraft accident report: Controlled flight into terrain, American Airlines Flight 965 Boeing 757-223, N651AA near Cali, Colombia December 20, 1995.* Santafe de Bogota, D.C. Colombia: Author.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* New York: Wiley.

Alexander, H. (1998). *Error Types and Related Error Detection Processes in the Aviation Domain.* Unpublished Master's Thesis, San Jose State University.

Andre T. and Wickens, C. D. (1995). When users want what's not best for them. *Ergonomics in Design, 3(4),* 10-14.

Barnett, B. and Wickens, C. D. (1986). *Non-optimality in the diagnosis of dynamic system states* (Technical Report CPL-86-8). Champaign, IL: University of Illinois, Cognitive Psychophysiology Laboratory, Department of Psychology.

Bennett, K. B., and Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34(5),* 513-533.

Billings, C. E. (1997). *Aviation automation: The search for a human centered approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Cha, W. and Funk, K. (1997). Recognizing pilots goals to facilitate agenda management. *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp.268-273). Columbus, OH: The Ohio State University.

Chandra, D. and Bussolari, S. R. (1991). An evaluation of flight path management automation in transport category aircraft. *Proceedings of the Sixth International Symposium on Aviation Psychology*, (pp.139-144). Columbus, OH: The Ohio State University.

Clark, H. H., And Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. M. Levine, and S. D. Teasley (Eds.) *Perspectives on socially shared cognition.* Washington, D.C.: American Psychological Association. pp. 127 - 149.

Clarke, A. A., and Smyth, M. G. G. (1993). A co-operative computer based on the principles of human co-operation. *International Journal of Man-Machine Studies, 38,* 3-22.

Conejo, R., and Wickens, C. D. (1997). *The effects of highlighting validity and feature type on air-to-ground target acquisition* performance (Technical Report ARL-97-11/NAWC-ONR-97-1) Savoy, IL: University of Illinois, Aviation Research Lab.

Coury, B. G., and Pietras, C. M. (1989). Alphanumeric and graphic displays for dynamic process control and monitoring. *Ergonomics, 32(11),* 1373-1389.

Coury, B. G., Boulette, M. D., and Smith, R. A. (1989). The impact of uncertainty on classification of multidimensional data with integral and separable displays. *Human Factors, 47,* 551-569.

Coury, B. G., and Semmel, R. D. (1996). Supervisory control and the design of intelligent user interfaces. In R. Parasuraman & M. Mouloua (Eds.) *Automation and human performance: Theory and applications.* (pp. 221 – 242). Mahwah, NJ: LEA.

Coury B. G., and Boulette, M. D.(1992). Time stress and the processing of visual displays. *Human Factors, 34(6),* 707-725.

Desauliniers, D. R., Gillan, D. J., and Rudisill, M. (1988). The effects of format in computer-based procedure displays. In *Proceedings of the Human Factors Society 32nd Annual Meeting.* (pp. 291 – 295). Santa Monica, CA: Human Factors Society.

Earlier control takeover might have averted A330 crash. (May 29, 1995). *Aviation Week & Space Technology,142(22),* 69-70.

Edland, A., and Svenson, O. (1993). Judgment and decision making under time pressure. In G. A. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.) *Decision Making in Action: Models and Methods.* (pp.27-40). Norwood, NJ: Ablex.

Egeth, H. E, and Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269-297.

Fox, M. (1988). An organizational view of distributed systems. In A. H. bond and L. Gasser (Eds.) *Readings in distributed artificial intelligence* (pp. 140-149). San Mateo, CA: Morgan Kaufman.

Gai, E. G., and Curry, R. E., (1976). A model of the human observer in failure detection tasks. *IEEE Transactions of Systems, Man and Cybernetics, SMC-6,* 85-91.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist, 25,* 350-358.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Garner, W.R., and Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing, *Cognitive Psychology, 1,* 225 - 241.

Gempler, K. S., and Wickens, C. D. (1998). *Display of predictor reliability on a cockpit display of traffic information.* (Technical Report ARL-98-6/Rockwell-98-1) Savoy, IL: University of Illinois, Aviation Research Lab.

Gillie, T., and Berry, D. (1994). Object displays and control of dynamic systems. *Ergonomics, 37(11),* 1885-1903.

Goldsmith, T. E., and Schvaneveldt, R. W. (1984). Facilitating multiple-cue judgments with integral information displays. In J. C. Thomas and M. L. Schneider (Eds.), *Human factors in computer systems* (pp. 243-270). Norwood, NJ: Ablex.

Groce, J.L., and Boucek, G. P. (1987). *Air transport crew tasking in an ATC data link environment.* (SAE paper 871764). Warrendale, PA: SAE.

Hahn, E. C. , and Hansman, R. J. (1992). *Experimental studies on the effect of automation on pilot situational awareness in the datalink ATC environment.* (SAE technical Report 851956). Warrendale, PA: SAE.

Hancock, P. A. (1993). On the future of hybrid human-machine systems. In J. S. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues* (pp. 61-86). Berlin: Springer-Verlag.

Hilburn, B., Jorna, P. G. A. M., and Parasuraman, R. (1995). The effect of advanced ATC automation on mental workload and monitoring performance: An empirical investigation in Dutch airspace. In *Proceedings of the 8th International Symposium on Aviation Psychology.* (pp. 387-391).Columbus, OH: The Ohio State University.

Hockey, G. J. R. (1984). Varieties of attentional state: The effect of the environment. In. R. S. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp.449-484). Orlando: Academic Press.

Hughes, R. L., Ginnett, R. C., and Curphy, G. J. (1996). Characteristics of the situation. In *Leadership: Enhancing the lessons of experience* (2nd ed., pp. 425 - 457). Chicago: Irwin.

Idaszak, J. R., and Hulin, C. L. (1989). *Active participation in highly automated systems: Turning the wrong stuff into the right stuff.* (ARL Report 89-7) Urbana-Champaign, IL: Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign.

Johnson, E. J., Payne, J. W., and Bettman, J. R. (1993). Adapting to time constraints. In O. Svenson and A. J. Maule (Eds.) *Time Pressure and Stress in Human Judgment and Decision Making.* (pp. 103-116). New York: Plenum Press.

Jones, P. M., Mitchell, C. M., and Rubin, K. S. (1990). Validation of intent inferencing by a model-based operator's associate. *International Journal of Man-Machine Studies, 33,* 177-202.

Jones, P.M., and Mitchell, C. M. (1995). Human-computer cooperative problem solving: Theory, design, and evaluation of an intelligent associate system. *IEEE Transactions on Systems, Man & Cybernetics, 25(7),* 1039-1053.

Jonides, J. and Yantis, S. (1988). Uniqueness of abrput visual onset in capturing attention. *Perception and Psychophysics, 43(4),* 346 – 354.

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. *Journal of Applied Psychology, 47(3),* 161-165.

Kahneman, D., and Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, and J. Beatty (Eds.), *Varieties of attention.* (pp.29-61). New York: Academic.

Kahneman, D., Treisman, A. M., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175-219.

Kerns, K. (1994). *Human factors in ATC/flight deck integration implications of data link simulation research.* (Mitre Report MP 94W0000098). McLean VA: Mitre.

Knox, C. E., and Scanlon, C. H. (1990). *Flight Tests Using Data Link for Air Traffic Control and Weather Information Exchange.* (SAE Paper 901888). Warrendale, PA. SAE.

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

Legge, G. E., Gu, Y., and Luebker, A. (1989). Efficiency of graphical perception. *Perception & Psychophysics, 46(4),* 365-374.

Liu, Y., Fuld, R., and Wickens, C. D. (1993). Monitoring behavior in manual and automated scheduling systems. *International Journal of Man-Machine Studies, 39(6),* 1015-1029.

Logsdon, E. W. (1996). *An Examination of Data Link Autoload and Message Length.* Unpublished Master's Thesis, San Jose State University.

Logsdon, E. W., Infield, S. E., Lozito, S., McGann, A., Mackintosh, M., and Possolo, A. (1995). Cockpit data link technology and flight crew communication procedures. In *Proceedings of the 8th International Symposium on Aviation Psychology.* (pp. 324 - 329). Columbus, OH: Ohio State University.

Lozito, S. , McGann, A., and Corker, K. (1993). Data link air traffic control and flight deck environments: Experiment in flight crew performance. In *Proceedings of the 7th International Symposium on Aviation Psychology*. (pp. 1009-1015). Columbus, OH: Ohio State University.

MacGregor, D., and Slovic, P. (1986). Graphic representation of judgmental information. *Human-Computer Interaction, 2,* 179-200.

Malone, T. W. (1988). Modeling coordination in organizations and markets. In A. H. bond and L. Gasser (Eds.) *Readings in distributed artificial intelligence* (pp. 151-158). San Mateo, CA: Morgan Kaufman.

Malone, T. and Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems? *Proceedings of the ACM Conference on Computer-Supported Cooperative Work,* October, 1990, 357 - 370.

Malone, T. and Crowston, K. (1993). The interdisciplinary study of coordination. *ACM Computing Surveys, 26(1),* 87-119.

May, P., Molloy, R., and Parasuraman, R. (1993). Effects of automation reliability and failure rate on monitoring performance in a multitask environment. Presentation at The Human Factors Society 37th Annual Meeting, Seattle, WA.

Milewski, A. E., and Lewis, S. H. (1997). Delegating to software agents. *International Journal of Human-Computer Studies, 46(4)*, 485-500.

Mosier, K. L., Skitka, L. J., and Korte, K. J. (1994). Cognitive and social issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *human performance in automated systems: Current research and trends* (pp. 191-197). Hillsdale, NJ: LEA.

Mosier, K. L., and Skitka, L. J. (1998). Automation bias and errors: Are teams better than individuals? In *Proceedings of the Human Factors and Ergonomics Society, 42nd Annual Meeting*. (pp. 201 – 205). Santa Monica, CA: Human Factors Society.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. Boff, L. Kaufman, & J. Thomas (Eds.). *Handbook of Perception and Performance: Vol 2. Cognitive Processes and Performance.* (pp. 40.1 - 40.51). New York: Wiley.

Muir, B. M. (1988). Trust between humans and machines, and the design of decision aids. In E. Hollnagel, G. Mancini & D. D. Woods, (Eds.), *Cognitive Engineering in Complex Dynamic Worlds* (pp. 71-84). London: Academic Press.

Muir, B. M. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. Unpublished doctoral dissertation, University of Toronto.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37(11),* 1905-1922.

Muir, B. M., and Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39(3),* 429-460.

Norman, D. A. (1990). The "problem" of automation: Inappropriate feedback and interaction, not "over-automation." *Philosophical Transactions of the Royal Society of London, B327,* 585-593.

Olson, W. A. and Sarter, N. B. (1998). "As long as I'm in control ...": Pilot preferences for and experiences with different approaches to automation management. In *Proceedings of the 4th Symposium on Human Interaction with Complex Systems.* (pp.63-72). IEEE.

Orasanu J. , and Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), *Decision Making in Action: Models and Methods.* (pp. 3-20). Norwood, NJ: Ablex.

Parasuraman, R., Molloy, R., and Singh, I. L. (1993). Performance consequences of automation-induced "complacency." *International Journal of Aviation Psychology, 3,* 1-23.

Pommerantz, J. R. (1986). Visual form perception: An overview. In H. C. Nusbaum and E. C. Schwab (Eds.), *Pattern Recognition by Humans and Machines* (Vol. 2, pp. 1 - 30). Orlando, FL: Academic Press.

Prevot, T., Palmer, E. and Crane, B. (1997). Flight crew support for automated negotiation of descent and arrival clearances. *Proceedings of the 7th International Symposium on Aviation Psychology.* (pp. 697 – 702). Columbus, OH: Ohio State University.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North-Holland.

Reason, J. T. (1990) *Human error.* New York: Cambridge University Press.

Rehmann, A. J. (1997). *Human Factors Recommendations for Airborne Controller-Pilot Data Link Communications (CPDLC) Systems: A Synthesis of Research Results and Literature.* (DOT/FAA/CT-TN97/6). Atlantic City, NJ: William J. Hughes Technical Center.

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49,* 95-112.

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

Riley, V. (1994). *Human Use of Automation.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications.* (pp. 19-35). Mahwah, NJ: LEA.

Ritchie, M. L. (1990). The missing (data) link in pilot navigation. In *IEEE PLANS '90: Position Location and Navigation Symposium Record.* (pp.634-637). New York: IEEE.

Roth, E. E., Bennett, K. B., and Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies, 27,* 479-525.

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes, 37,* 83-92.

Rouse, W. B., and Morris, N. M. (1987). Conceptual design of a human error tolerant interface for complex engineering systems. *Automatica, 23(2),* 231-235.

RTCA (1995). *Report of the RTCA Board of Directors' Select Committee on Free Flight.* Washingon DC: Author.

Saavedra, R., Earley, P. C., and Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, (78)*, 61-72.

Sanderson, P.M. (1989). The human planning and scheduling role in advanced manufacturing systems: an emerging human factors role. *Human Factors, 31,* 635-666.

Sanderson, P. M., Flach, J. M., Buttieg, M. A., and Casey, E. J. (1989). Object displays do not always support better integrated task performance. *Human Factors, 31,* 183-198.

Sanderson, P. M., Haskell, I., and Flach, J. M. (1992). The complex role of perceptual and organization in visual display design theory. *Ergonomics, 35(1),* 1199 - 1219.

Sarter, N. B. (1995). "Knowing when to look where": Attention allocation on advanced automated flight decks. In R. S. Jensen (Ed.), *Proceedings of the 8th International Symposium on Aviation Psychology* (pp. 239 –242). Columbus OH: Ohio State University.

Sarter N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *The International Journal of Aviation Psychology, 2(4),* 303-321.

Sarter N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *International Journal of Aviation Psychology, 4(1),* 1-28.

Sarter N. B., and Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37(1),* 5-19.

Sarter, N. B., and Woods, D. D. (1997). Teamplay with a powerful and independent agent: A corpus of operational experiences and automation surprises on the airbus A-320. *Human Factors, 39(4),* 553-569.

Sarter, N. B., Woods, D. D., and Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1926 – 1943). New York: John Wiley and Sons.

Sarter, N. B., and Woods, D. D. (submitted). Observations of teamplay with a powerful and independent agent in a full mission simulation study. *Human Factors.*

Schwartz, D. R., and Howell, W. C. (1985). Optional stopping performance under graphic and numeric CRT formatting. *Human Factors, 27(4),* 433-444.

Sekigawa, E., and Mecham, M. (1996, July 29). Pilots, A300 systems cited in Nagoya crash. *Aviation Week & Space Technology, 145(5),* 36-37.

Sellen, A. J. (1994). Detection of everyday errors. *Applied Psychology: An International Review, 43(4),* 475-498

Sheridan, T. (1997). Supervisory control. In G. Salvendy (Ed.) *Handbook of human factors and ergonomics.* (pp. 1295 – 1327). New York: John Wiley and Sons.

Sheridan, T. B. & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators.* Cambridge, MA: MIT Press.

Silverman, B. G. (1992). Human-computer collaboration. *Human-Computer Interaction, 7,* 165-196.

Sorkin, R. D., Mabry, T. R., Weldon, M. S., and Elvers, G. (1991). Integration of information from multiple element displays. *Organizational Behavior and Decision Processes, 49,* 167-187.

Spence, R. and Parr, M. (1991). Cognitive assessment of alternatives. *Interacting with Computers, 3(3),* 270-282.

Steeb, R., Cammarata, D., Hayes-Roth, F., Thorndyke, P., and Wesson, R. (1988). Architectures for distributed air-traffic control. In A. H. bond and L. Gasser (Eds.) *Readings in distributed artificial intelligence* (pp. 90-101). San Mateo, CA: Morgan Kaufman.

Suchman, L. A. (1987). *Plans and situated actions.* Cambridge: Cambridge University Press.

Symon, G., Long, K., and Ellis, J. (1996). The coordination of work activities: cooperation and conflict in a hospital context. *Computer Supported Cooperative Work: The Journal of Collaborative Computing, 5,* 1-31.

Van Gent, R.N.H.W. (1995). *Human Factors Issues with airborne data link: towards increased crew acceptance for both enroute and terminal flight operations.* (NLR report TP 95666 L). Amsterdam: NLR.

Waller, M. C. (1992). *Flight Deck Benefits of Integrated Data Link Communication.* (NASA TP 3219). Hampton, VA: NASA Langley Research Center.

Wickens, C. (1992). Process control and automation. In *Engineering Psychology and Human Performance.* (pp. 504-551).New York: Harper-Collins.

Wickens, C. D. and Kessel, C. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-9,* 24-34.

Wickens, C. D. and Kessel, C. (1980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology, Human Perception and Performance, 6,* 564-577.

Wickens, C. D. and Scott, B. D. (1983). *A comparison of verbal and graphical information presentation in complex information integration decision task* (Report EPL-83-1/ONR-83-1). Urbana, IL: Engineering Psychology Lab, University of Illinois.

Wickens, C. D., Stokes, A. Barnett, B., and Hyman, F. (1993). The effects of stress on pilot judgment in a MIDIS simulator. In O. Svenson and A. J. Maule (Eds.) *Time Pressure and Stress in Human Judgment and Decision Making.* (pp. 271-292). New York: Plenum Press.

Wickens, C. D. and Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37(3),* 473-494.

Wiener, E. L. (1985). *Cockpit automation: In need of a philosophy* (SAE Technical Report 851956). Warrendale, PA: SAE.

Wiener, E. L. (1989) *Human factors of advanced technology ("glass cockpit") transport aircraft* (NASA Technical Report 117528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23*, 955-1011.

Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning.* Reading, MA: Addison-Wesley.

Woods, D. D. (1995). Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In J. Flach, P. Hancock, J. Caird, & K. Vincente (Eds.), *Global perspectives on the ecology of human-machine systems.* (pp. 157 – 188). Hillsdale, NJ: LEA.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications.* (pp. 3 - 18). Mahwah, NJ: LEA.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers, and hindsight.* (CSERIAC State-of-the-art Report SOAR 94-01). Wright-Patterson AFB, OH: CSERIAC.

Wright, P. (1974). The harassed decision maker: Time pressure, distraction and the use of evidence. *Journal of Applied Psychology, 59,* 555-561.

Zachary, W. W., and Robertson, S. P. (1990). Introduction. In S. P. Robertson, W. W. Zachary, and J. B. Black (Eds.) *Cognition, computing and cooperation.* (pp.1 – 21). Norwood, NJ: Ablex.

Appendices

Appendix A. *Scenario Information*

Inappropriate Goal Conflict – Routine (Vector 1 mile inside final approach fix)

Routing: LOCKE9 arrival to SFO

Initial altitude: 12,000

Initial Speed: 265

Initial heading: 287

Initial position: LOCKE

Initial routing (list of waypoints) LOCKE, UPEND, ILS RWY 19

Clearances

Initial Clearance: "Cleared to KSFO via LOCKE9 arrival, maintain 10,000"

Clearance 1: "Descend and maintain 5,000, speed 250 kts"

Location: 25 before UPEND

Clearance 2: "turn left heading 240, vectors ILS 28R"

Location: 20 before UPEND

Clearance 3: "Descend and maintain 2,000, contact approach 126.95"

Location: 5 miles later

Clearance 4: "**Urgent** – Turn left heading 190, maintain speed 230 kts"

Location: 20 out of airport

Clearance 5: "Fly heading 220, speed your discretion, maintain 2,000 until established, cleared ILS runway 28R" (Conflict – heading 220 puts aircraft approximately 1 mile inside of the final approach fix)

Location: 10 from final approach course

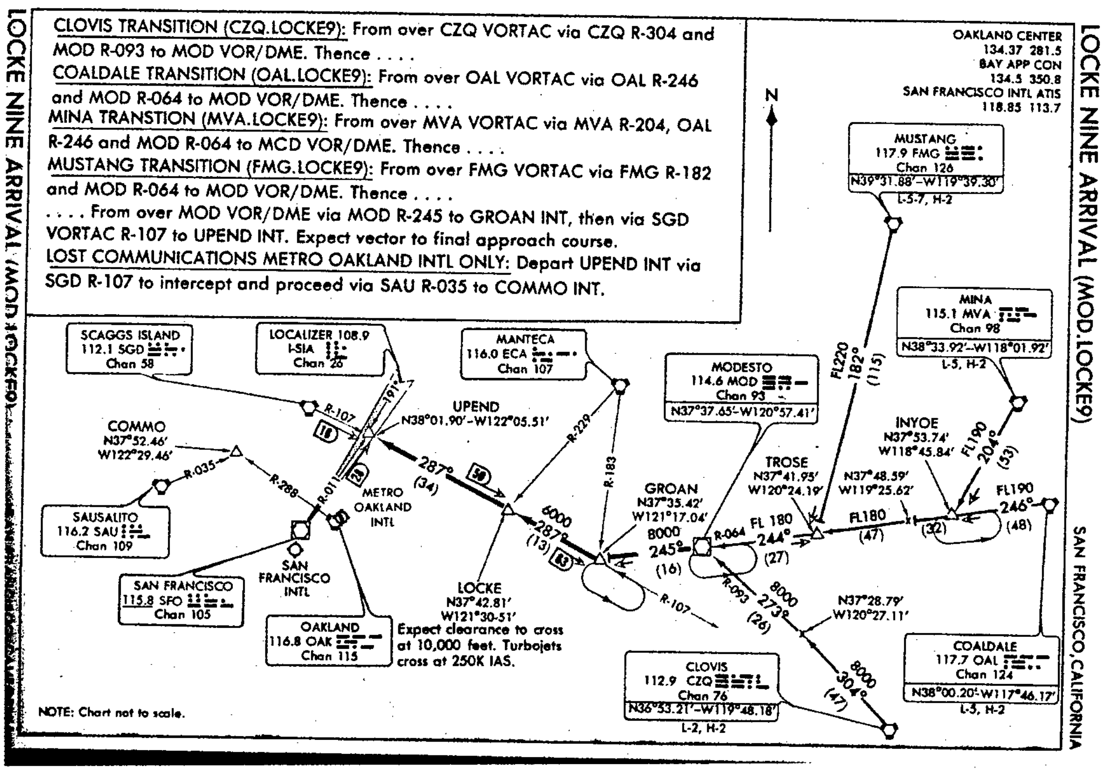


Figure A1. Routing for inappropriate goal conflict – routine.

Inappropriate Goal Conflict – Urgent (Speed restriction is too fast)

Routing: DOWNE2 arrival to LAX

Initial altitude: 10,000

Initial Speed: 250

Initial heading: 248

Initial position: CIVET (34.034N 117.39W)

Initial routing (list of waypoints): CIVET BASET DOWNE ILS 25

Clearances:

Initial Clearance: "Cleared to KLAX via DOWNE2 arrival, maintain 10,000"

Clearance 1: "Descend and maintain 6,000, contact approach 124.5"

Location: At CIVET

Clearance 2: "Descend and maintain 3,500, maintain speed 240 kts"

Location: 18 out of BASET

Clearance 3: "Fly heading 230, vectors ILS 25L, slow to 230 kts"

Location: 12 miles out of DOWNE

Clearance 4: "**Urgent** - fly heading 265, maintain 2,100 until established, cleared ILS 25L, maintain 220 kts until LIMMA" (conflict – 220 is too fast since LIMMA is only 5 miles from runway)

Location: BASET

Clearance 5: "Contact tower 120.95"

Location: near DOWNE.

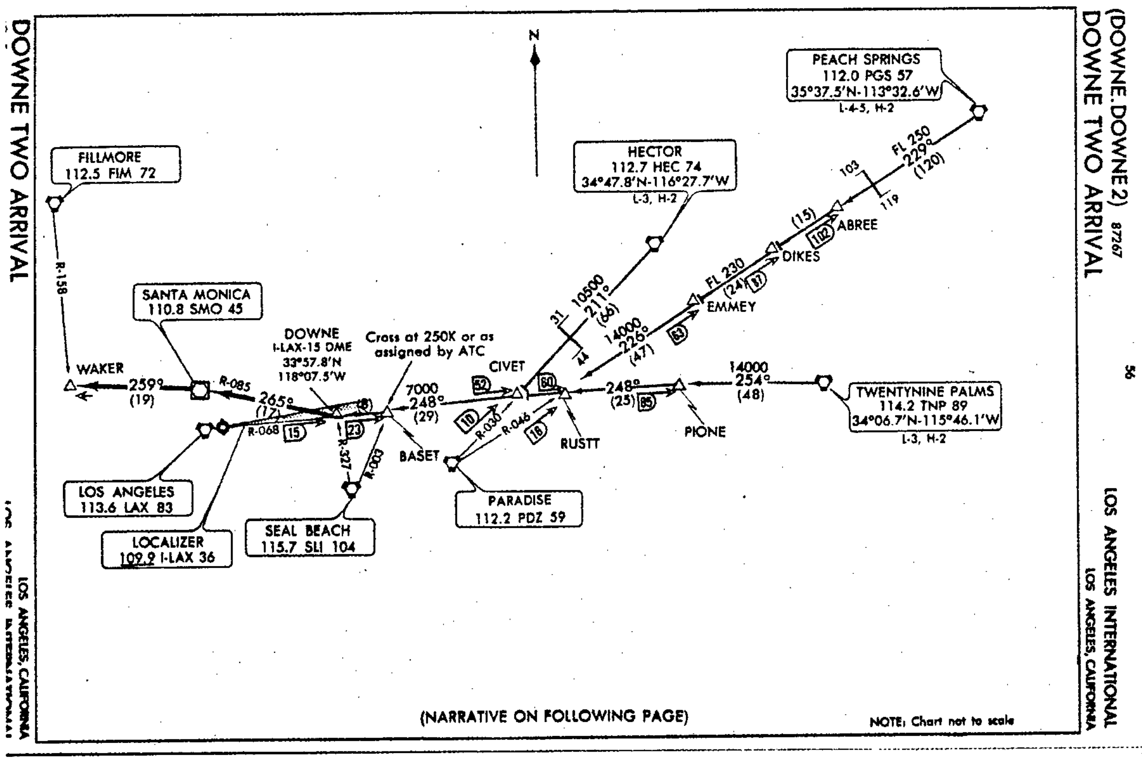


Figure A2. Routing for inappropriate goal conflict – urgent.

Impossible Goal Conflict – Routine (Slow to speed above current speed)

Routing: AGGIT2 arrival to HOU

Initial altitude: FL 290

Initial Speed: 292

Initial heading: 049

Initial position: 20 past FLAKY (28.5068N 96.3332W)

Initial routing (list of waypoints) FLAKY AGGIT TIDDY

Clearances

Initial Clearance: "Cleared to KHOU via AGGIT2 arrival, maintain FL 290"

Clearance 1: **"Urgent**, For traffic, fly heading 010, descend and maintain FL 240"

Location: at X

Clearance 2: "Proceed direct AGGIT, AGGIT2 arrival, continue descent to 17,000, contact center 127.95"

Location: 40 out of AGGIT

Clearance 3: "For spacing, fly heading 030, descend and maintain 12,000, increase speed to 250" (conflict – current speed is above 250)

Location: 30 out of AGGIT

Clearance 4: "Direct TIDDY, AGGIT2 arrival, resume normal speed, cross TIDDY at and maintain 5,000"

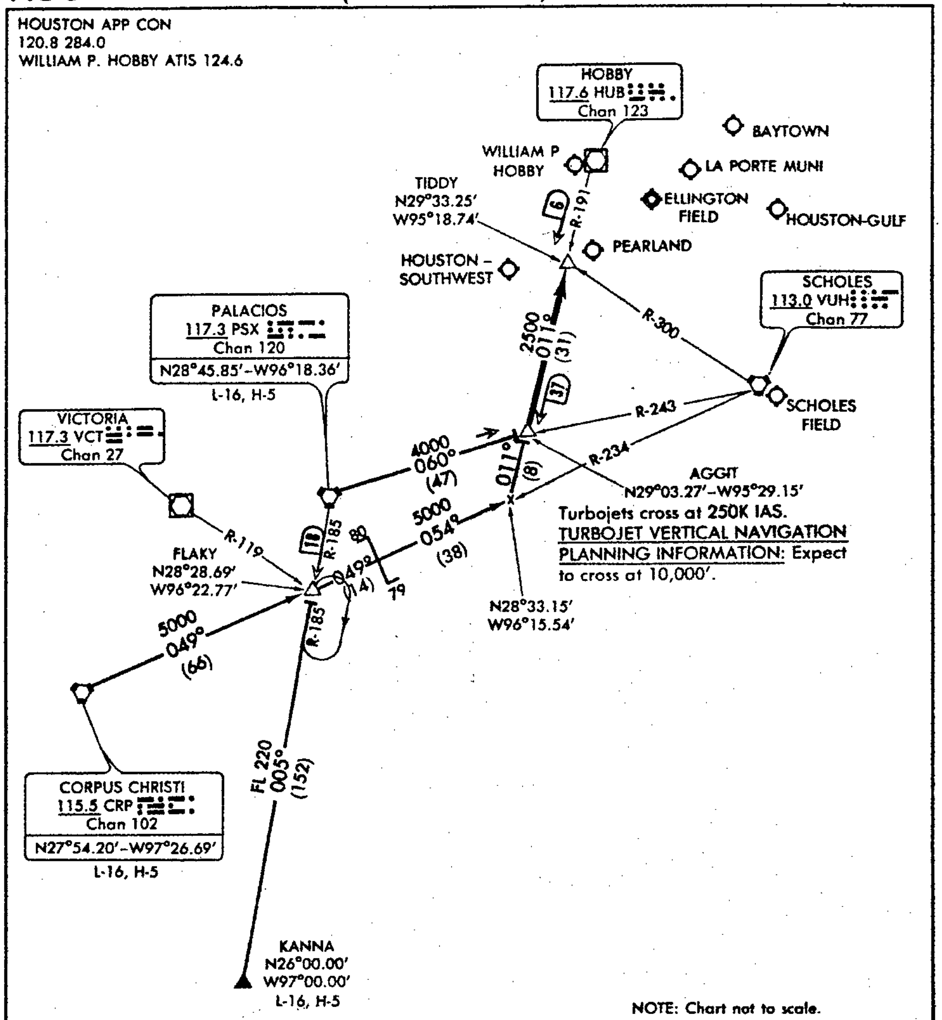
Location: 45 out of TIDDY (21 out of AGGIT)

Clearance 5: "Fly heading 050, stop descent at 8,000"

Location: 35 out of TIDDY

87323

AGGIT TWO ARRIVAL (AGGIT.AGGIT2) HOUSTON, TEXAS



CORPUS CHRISTI TRANSTION (CRP.AGGIT2): From over CRP VORTAC via CRP R-049, VUH R-234, and HUB R-191 to AGGIT INT. Thence . . . .
KANNA TRANSITION (KANNA.AGGIT2): From over KANNA INT via PSX R-185 to FLAKY INT then via CRP R-049, VUH R-234, and HUB R-191 to AGGIT INT. Thence . . . .
PALACIOS TRANSITION (PSX.AGGIT2): From over PSX VORTAC via PSX R-060 to AGGIT INT. Thence . . . .
. . . . From over AGGIT INT via HUB R-191 to TIDDY INT. MEA 2500. Expect vectors to final approach course.

AGGIT TWO ARRIVAL (AGGIT.AGGIT2) HOUSTON, TEXAS

3

Figure A3. Routing for impossible goal conflict – routine.

Impossible Goal Conflict – Urgent (Descend to altitude above current altitude)

Routing: BLUFI 4 arrival to MIA (on heading120)

Initial altitude: 27,000

Initial Speed: 305

Initial heading: 120

Initial position: 15 NW of BLUFI (27.0769N 79.9857W)

Initial routing (list of waypoints) BLUFI HONOE BSY

Clearances

Initial Clearance: "Cleared to KMIA via heading 120, expect BLUFI4 arrival, maintain FL 270"

Clearance 1: "Cleared direct BLUFI, BLUFI4 arrival, descend and maintain FL 230"

Location: 12 NW of BLUFI

Clearance 2: "Descend and maintain FL 190, contact Miami center freq 123.45"

Location: 5 out of BLUFI

Clearance 3: "Descend and maintain 15,000, cross HONOE at 15,000, 300 kts"

Location: 18 out of HONOE

Clearance 4: "**Urgent** For traffic, descend and maintain 17,000, fly heading 180" (conflict – 17,000 is above current altitude)

Location: level at 15,000

Clearance 5: "Fly heading 210, descend and maintain10,000, contact Miami approach 128.95"

Location: 40 out of BSY

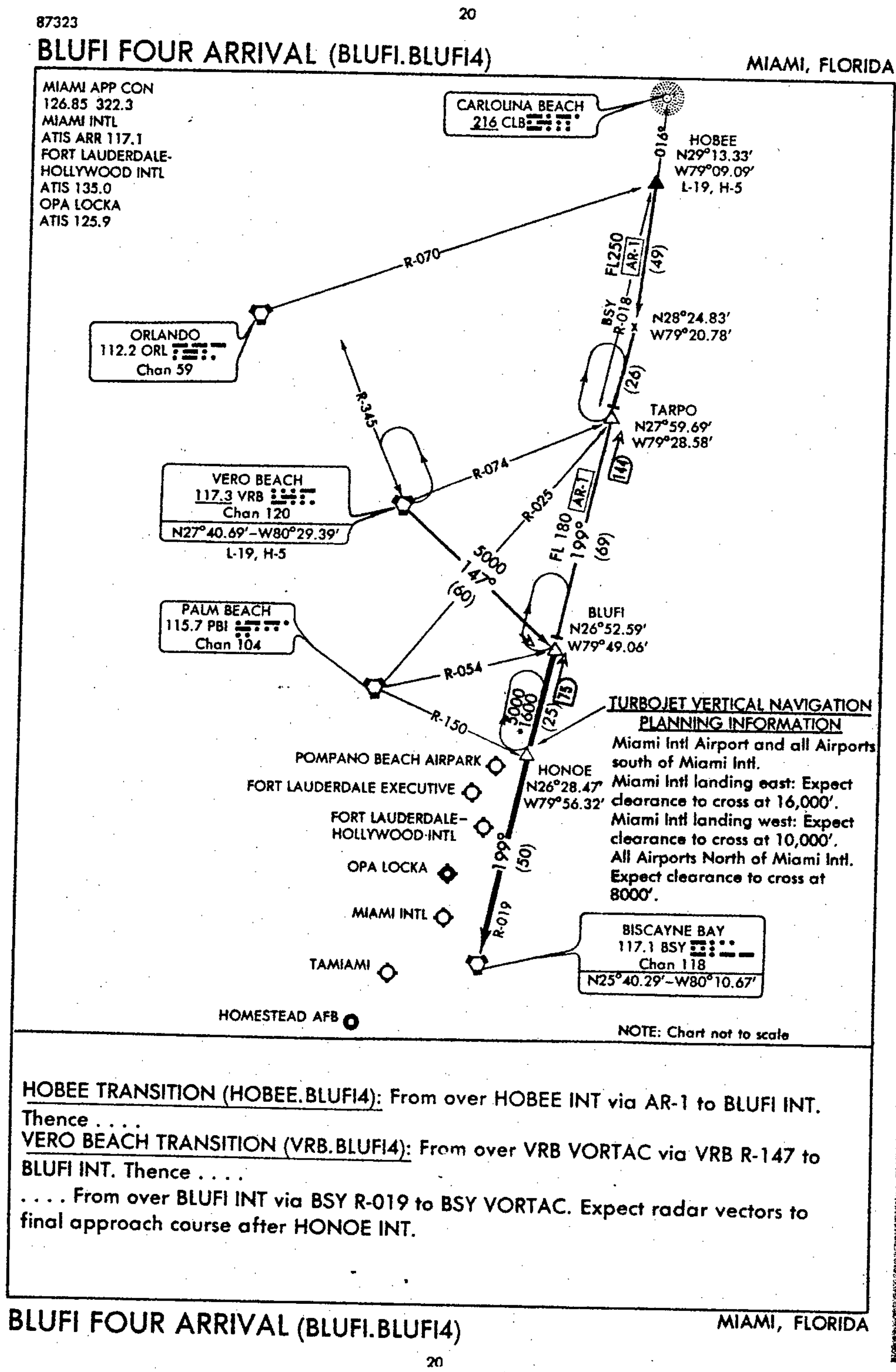


Figure A4. Routing for impossible goal conflict – urgent.

Implementation Conflict: Automation Does Less – Routine (Speed restriction does not propagate from cruise to descent)

Routing: WARRD2 arrival to EWR

Initial altitude: 290

Initial Speed: 292

Initial heading: 044

Initial position: ENO (39.061N 75.5163W)

Initial routing (list of waypoints) ENO DAVYS HOLEY RBV WARRD

Clearances

Initial Clearance: "Cleared to KEWR via WARRD2 arrival, maintain FL 290"

Clearance 1: **"**For spacing, maintain speed 280 until advised, contact center 128.45"

Location: ENO

Clearance 2: "Descend and maintain 17,000, cross DAVYS at 17,000" (conflict - when pilot engages VNAV, speed is no longer set to 280)

Location: 42 north of DAVYS

Clearance 3: **"Urgent** – For traffic, fly heading 080, resume normal speed, contact Newark approach 126.85"

Location: 25 out of DAVYS

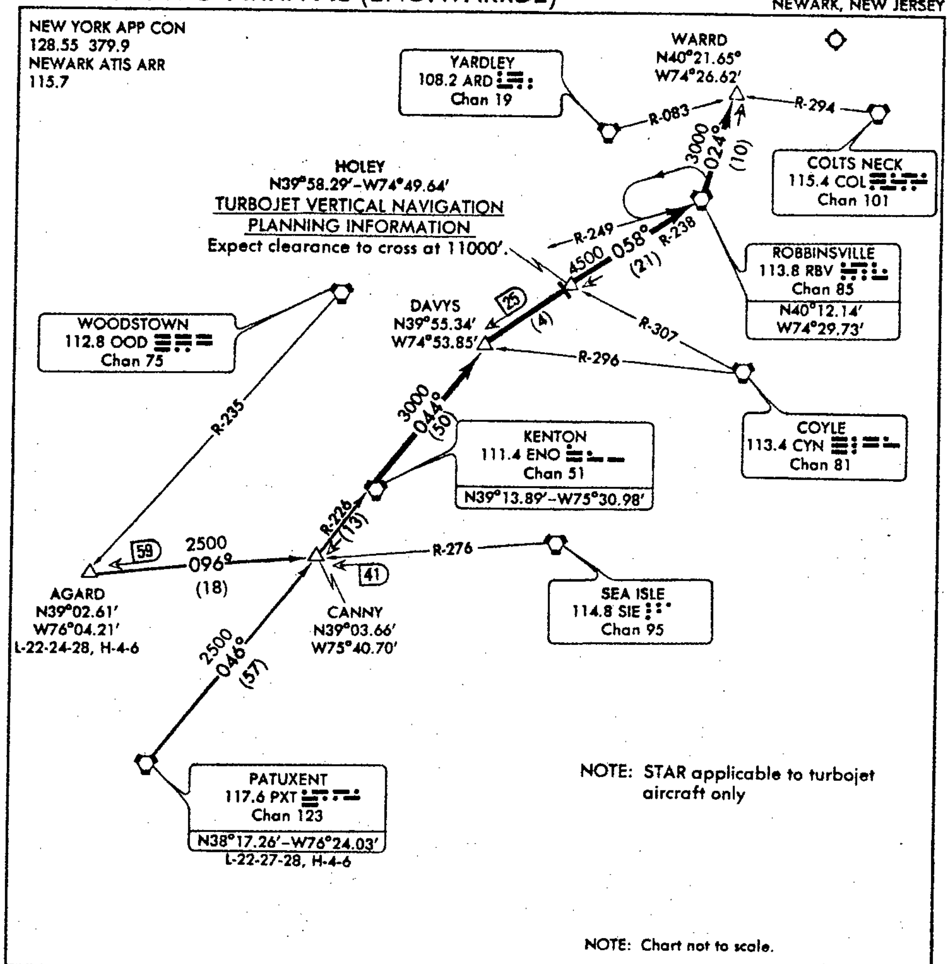
Clearance 4: "Cleared direct RBV, resume WARRD2 arrival, descend and maintain 10,000;"

Location: 20 out of DAVYS

Clearance 5: "Descend and maintain 8,000, Cross RBV at 8,000, 230 kts"

Location: 35 out of RBV

WARRD TWO ARRIVAL (ENO.WARRD2) NEWARK INTL
NEWARK, NEW JERSEY



AGARD TRANSITION (AGARD.WARRD2): From over AGARD INT via SIE R-276 and ENO R-226 to ENO VORTAC. Thence . . . .

PATUXENT TRANSITION (PXT.WARRD2): From over PXT VORTAC via PXT R-046 and ENO R-226 to ENO VORTAC. Thence . . . .

. . . . From ENO VORTAC via ENO R-044 and RBV R-238 to HOLEY INT, then via RBV R-238 to RBV VORTAC, then via RBV R-024 to WARRD INT. Expect radar vectors to final approach course after WARRD INT.

WARRD TWO ARRIVAL (ENO.WARRD2) NEWARK, NEW JERSEY
NEWARK INTL

171

Figure A5. Routing for implementation conflict – automation does less – routine.

Implementation Conflict: Automation Does Less – Urgent (Along track waypoint does not load because it is beyond subsequent waypoint)

Routing: ESL then ARMEL1 arrival to DCA

Initial altitude: FL 210

Initial Speed: 280

Initial heading: 107

Initial position: ESL (39.23N 78.99W)

Initial routing (list of waypoints) ESL DRUZZ TRIXY AML

Clearances

Initial Clearance: "Cleared to KDCA via ARMEL1 arrival, maintain FL 210"

Clearance 1: "Descend and maintain 15,000, cross DRUZZ at 15,000"

Location: 24 out of DRUZZ

Clearance 2: "**Urgent** Descend and maintain 11,000, cross DRUZZ at 15,000, 280 knots, continue to cross 15 east of DRUZZ at 12,000," (conflict – since 15 east of DRUZZ is past TRIXY, it will not load into FMC as DRUZZ +15)

Location: 12 out of DRUZZ

Clearance 3: "Contact Washington approach 126.45"

Location: at DRUZZ

Clearance 4: "Vectors for sequencing Fly heading 150, maintain speed 250 kts"

Location: past TRIXY

Clearance5: "Fly heading 190, descend and maintain 8,000"

Location: 25 out of AML

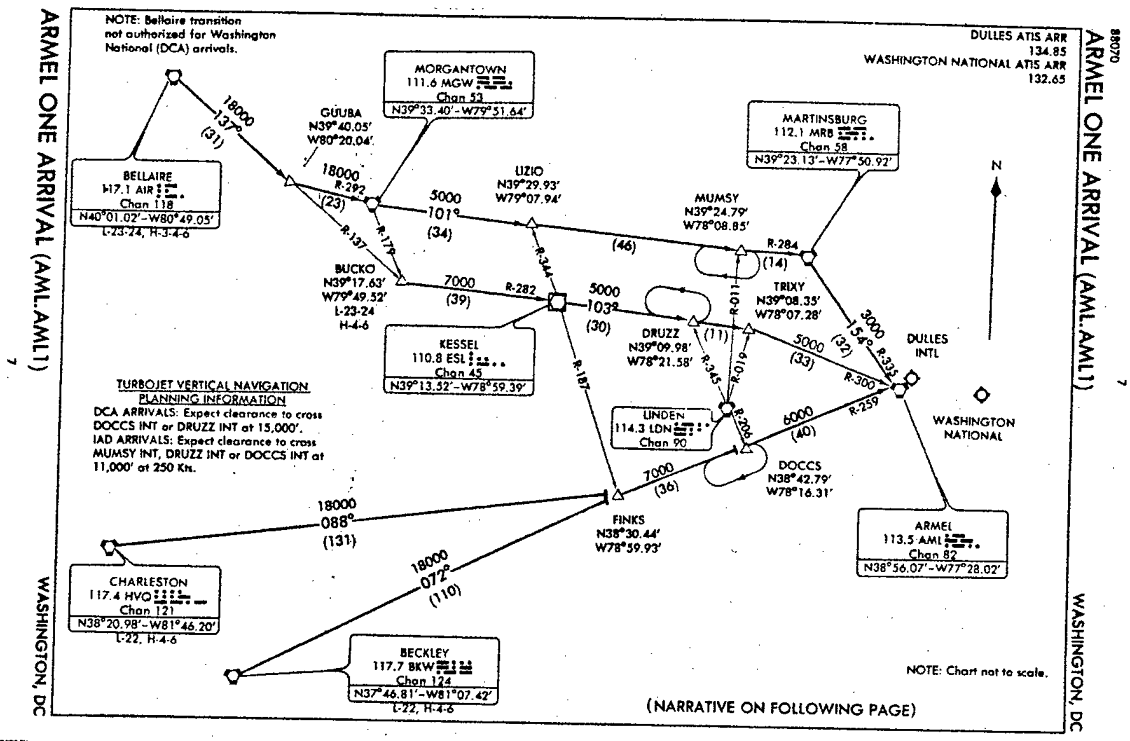


Figure A6. Routing for implementation conflict – automation does less – urgent.

Implementation Conflict: Automation Does More – Routine (Change in routing deletes vertical profile)

Routing: AQN9 arrival to DFW

Initial altitude: 21000

Initial Speed: 305

Initial heading: 075

Initial position: DANCR (32.401N 98.0553W)

Initial routing (list of waypoints) COTTN DANCR AQN MARKM BRYAR FLATO CREEK

Clearances

Initial Clearance: "Cleared to KDFW via AQN9 arrival, maintain FL 210"

Clearance 1: **"Urgent** For traffic, fly heading 110, descend and maintain FL 190, maintain 280 kts"

Location: immediate

Clearance 2: "Proceed direct AQN then AQN9 arrival, assigned Runway 17R, descend and maintain 15,000, speed your discretion"

Location: 12 out of AQN

Clearance 3: "Descend and maintain 10,000, cross BRYAR at 12,000"

Location: 5 before AQN

Clearance 4: "Change assigned runway to 17L, contact approach 123.45" (conflict – deletes 12,000 restriction at BRYAR in FMC)

Location: Past AQN

Clearance 5: "Continue descent to 5,000,"

Location: past BRYAR

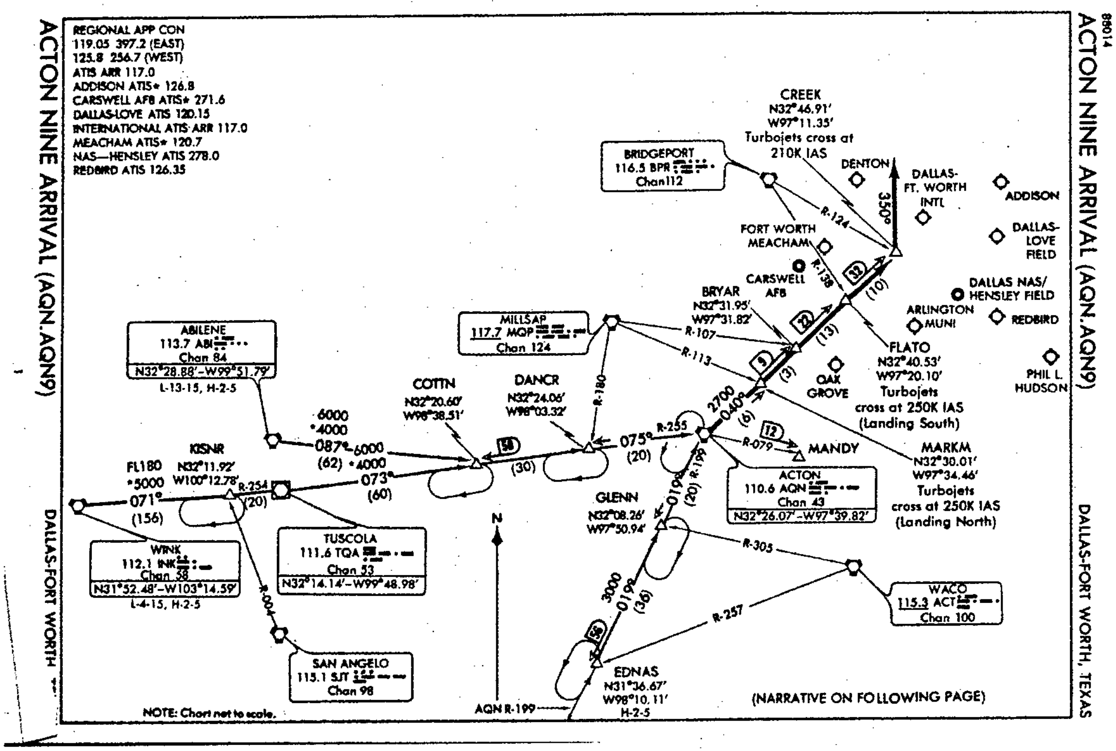


Figure A7. Routing for implementation conflict – automation does more – routine.

Implementation Conflict: Automation Does More – Urgent (Deleting previously given airspeed restriction also deletes an altitude restriction)

Routing: La Grange six arrival into Atlanta

Initial altitude: 25,000

Initial Speed: 305

Initial heading: 048°

Initial position: x (32.8003N 85.6637W)

Initial routing (list of waypoints) TIMMY, LGC, HONIE, TIROE, ATL

Clearances

Initial Clearance: "Cleared to KATL via LGC6 arrival, maintain FL 250"

Clearance 1: " For spacing, fly heading 020, descend and maintain FL 230, slow to 280 kts"

Location: 4 miles from TIMMY

Clearance 2: "Cleared direct LGC, LGC6 arrival, speed your discretion, descend and maintain FL210"

Location: 18 miles from LGC

Clearance 3: "Continue descent to and maintain 10,000, cross TIROE at 12,000, 270 kts"

Location: 8out of LGC

Clearance 4: "Contact Atlanta approach, 123.75"

Location: 8 out of HONIE

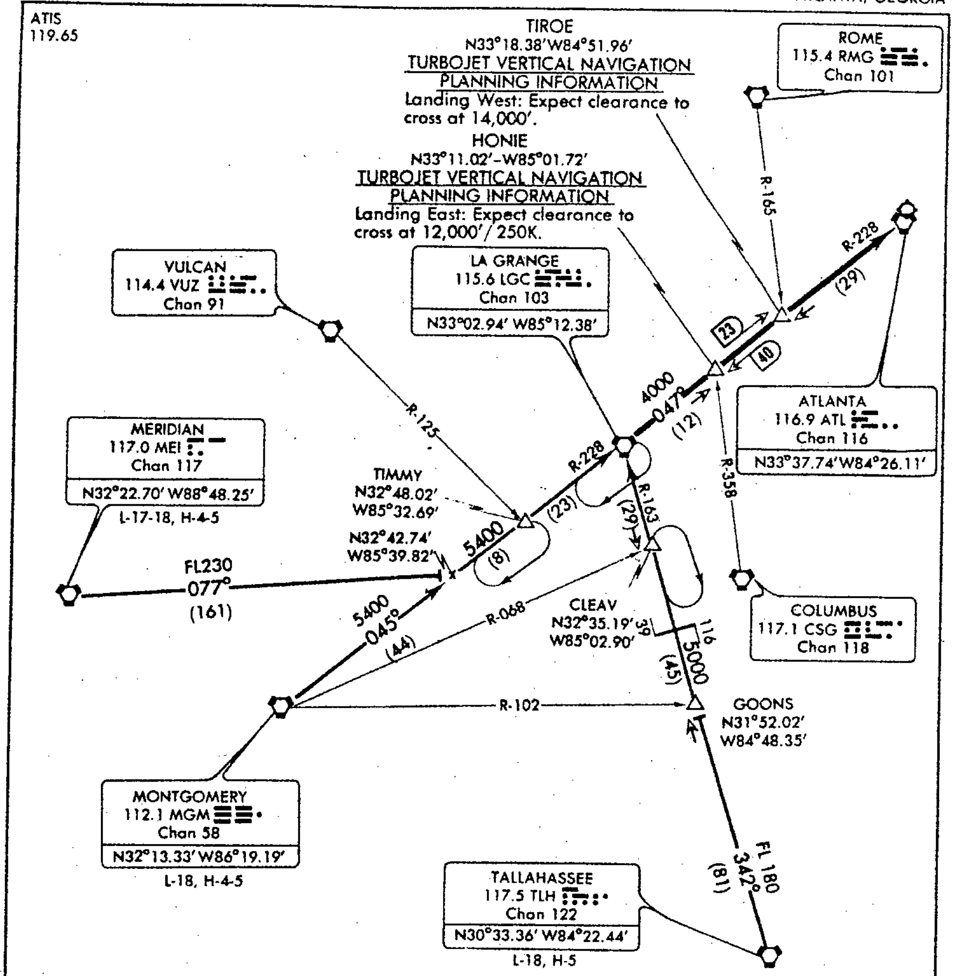
Clearance 5: "**Urgent** Increase rate of descent, speed your discretion at TIROE"(conflict – deletes 12,000 restriction at TIROE)

Location: 8 before TIROE

(LGC.LGC6)87323

LA GRANGE SIX ARRIVAL

THE WILLIAM B. HARTSFIELD ATLANTA INTL
ATLANTA, GEORGIA



NOTE: Chart not to scale.

MERIDIAN TRANSITION (MEI.LGC6): From over MEI VORTAC via MEI R-077 and LGC R-228 to LGC VORTAC. Thence . . . .

MONTGOMERY TRANSITION (MGM.LGC6): From over MGM VORTAC via MGM R-045 to LGC VORTAC. Thence . . . .

TALLAHASSEE TRANSITION (TLH.LGC6): From over TLH VORTAC via TLH R-342/116 NM and LGC R-163/39 NM to LGC VORTAC. Thence . . . .

. . . . from over LGC VORTAC via ATL R-228 to ATL VORTAC. Expect radar vector to final approach course after TIROE INT.

LA GRANGE SIX ARRIVAL
(LGC.LGC6)

91

ATLANTA, GEORGIA
THE WILLIAM B. HARTSFIELD ATLANTA INTL

Figure A8. Routing for implementation conflict – automation does more – urgent.

Appendix B. *Trust Rating Sheet*

Trust Rating Sheet

Please place a mark indicating your level of trust in items mentioned in each statement. Even though each scale is divided by a set of scale markings, you may place a mark at any point along the line. Remember there are no right or wrong answers, we are interested in your impressions and feelings.

1. I trust air traffic controllers to provide safe and acceptable clearances.

Not at all — Completely

2. I trust data link systems to load a set of targets to the MCP and FMC CDU in an acceptable manner based on a pending data link clearance.

Not at all — Completely

3. I trust the FMS to implement the targets provided by data link systems in an acceptable manner.

Not at all — Completely

Appendix C. *Table of Conflict Detection Stages*

Conflicts detected by conflict type and condition

	No Gating				Gating/Text								Gating/Graphic								
Conflict Detection Stage/Activity	G1 R	G1 U	G2 R	G2 U	G1 R	G1 U	G2 R	G2 U	I1 R	I1 U	I2 R	I2 U	G1 R	G1 U	G2 R	G2 U	I1 R	I1 U	I2 R	I2 U	
Pre-Consent Evaluation	2	1	7	7	2	1	6	2					2	1	6	2					39
Post-Consent Confirmation	1	2	1	2	3	1		5	3	8	2	3	4	1	2	7	2	8	1	1	55
Subsequent Monitoring	3			1	1	5			2	2			2	2			1	1	2	1	23
Forcing Function		2			2	1							1	2							8
total	6	5	8	10	8	8	6	7	5	10	2	3	9	6	8	9	3	9	3	2	125

G1 – INAPPROPRIATE GOAL CONFLICT
G2 – IMPOSSIBLE GOAL CONFLICT
I1 – IMPLEMENTATION – LESS
I2 – IMPLEMENTATION – MORE

R – Routine
U – Urgent

Vita

WESLEY ALLAN OLSON

Home: 3725 Cricklewood Dr
Montgomery, AL 36109
(334) 396-3405

Office: Air Command and Staff College
Maxwell AFB, AL
(334) 953-6794

EDUCATION

University of Illinois at Urbana-Champaign — Aug 96 – Jul 99
Urbana-Champaign, Illinois
Ph.D. in Psychology (October 1999)

University of Illinois at Urbana-Champaign — Aug 85 – Jan 87
Urbana-Champaign, Illinois
M.S. in Psychology (Jan 87)

United States Air Force Academy — Jun 81 – May 85
Colorado Springs, Colorado
B.S. (May 85)
- Distinguished Graduate
- Outstanding Cadet in Behavioral Science

EXPERIENCE

University of Illinois at Urbana-Champaign — Aug 96 – Jul 99
Urbana-Champaign, Illinois

Graduate student. Participated in research at the Aviation Research Laboratory at the University of Illinois. Conducted a line of research examining the human-machine coordination implications of various automation management strategies. Designed, administered, and analyzed and reported on automation survey data. Formulated experimental design, displays and procedures, coordinated with major US airline, collected and analyzed data, wrote reports on experimental research into human-machine coordination, presented results at professional meetings.

United States Air Force Academy — Oct 93 – Aug 96
Colorado Springs, Colorado

Assistant professor in Behavioral Science Department, flight instructor. Taught courses in introductory human factors, application of human factors design principles, aviation psychology, and leadership. Developed curriculum and course materials. Course director for introductory human factors and aviation psychology courses; supervised other instructors and oversaw test development. Served as academic advisor. Managed department budget. Participated in leadership development activities and organizational interventions for organizations within the Academy as well as in the Department of Defense. Taught human factors principles of display design at US Air Force Test Pilot school. Instructed cadets and trained qualified pilots in the TG7A motorized glider as part of the glider training program. Conducted research in crew resource management and cooperative education.

Travis Air Force Base Jun 91 – Oct 93
Fairfield, California

C-5 instructor pilot and Wing Airlift Director. Flew cargo and passengers to a variety of worldwide locations in support of military and humanitarian missions. Managed flight crew of 8 – 18 personnel. Provided flight instruction and recurrent training. As squadron chief of training re-instituted training programs following the Gulf War. As wing airlift director scheduled and coordinated all airlift missions departing Travis AFB.

Andrews Air Force Base Mar 88 – Jun 91
Camp Springs, Maryland

C-21 flight examiner and assistant chief of standardization/evaluation. Flew high ranking military and civilian dignitaries to a variety of locations in the US. Administered flight evaluations and provided flight training. Tracked and ensured compliance with applicable training and operating regulations. As mobility officer was responsible for developing and executing deployment of two aircraft and 15 personnel to Europe.

Columbus Air Force Base Jan 87 – Mar 88
Columbus, Mississippi

Student pilot. Completed 200 hour of flight instruction in T-37 and T-38 aircraft. Distinguished graduate.

University of Illinois at Urbana-Champaign Aug 85 – Jan 87
Urbana-Champaign, Illinois

Research assistant in Department of Psychology. Conducted research on manual control performance with various types of visual feedback. Developed experimental design, collected data, analyzed results, presented results at professional meeting.

United States Air Force Academy Jul 84 – Aug 85
Colorado Springs, Colorado

Conducted research in the display of attitude information. Analyzed data, wrote report, presented results at professional meeting.

Edwards Air Force Base May 84 – Jul 84
Edwards, California

Test and evaluation assistant in the Human Factors Branch at the US Air Force Flight test center. Assisted in developing and conducting evaluation of night air refueling capability for the B-1 bomber.

PUBLICATIONS

Olson, W. A., and Sarter, N. B. (In press). Supporting informed consent in human-machine Collaboration: The role of conflict type, time pressure, and display design. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors and Ergonomics Society.

Olson, W. A., and Sarter, N. B. (Submitted). Pilot preferences for and operational experiences with automation management strategies and implementations. *International Journal of Aviation Psychology.*

Olson, W. A. and Sarter, N. B. (1998). "As long as I'm in control ...": Pilot preferences for and experiences with different approaches to automation management. In *Proceedings of the 4th Symposium on Human Interaction with Complex Systems.*(pp.63-72). IEEE.

McKinney, E., Barker, J., Woody, J., Garvin, J. D., and Olson, W. A. (1997). Recent experience and tanker crew performance: The impact of familiarity on crew coordination. In *Proceedings of the Ninth Symposium on Aviation Psychology* (pp. 586 – 590). Columbus, OH: The Ohio State University.

Harwood, K., Smith, K., Olson, W. A., and Hancock, P. (1994). Shared decision making in the National Airspace System: Flightdeck-ATC integration. In *Proceedings of the 14th Symposium on Applied Behavioral Sciences* (pp. 188 – 193). Colorado Springs, CO: United States Air Force Academy.

Flach, J. M., Hagen, B. A., O'Brien, D., Olson, W. A. (1990). Alternative displays for discrete movement control. *Human Factors. 32 (6)*, 685-695.

Olson, W. A., and Flach, J. M. (1986). An exploration of a Fitts' Law task on a logarithmically scaled screen. In *Proceedings of the Twenty-Second Annual Conference on Manual Control* (pp. 165-173). Dayton, OH: Air Force Wright Aeronautical Laboratories.

FLIGHT RATINGS

FAA - Air Transport Pilot – Lear type rating
USAF – Senior Pilot (3000 hours)
Aircraft flown – TG7A, C-5, C-21, T-37, T-38