Iowa State University
Digital Repository

2014

Towards a characterization of information automation systems on the flight deck

Rachel Feddersen Dudley
Iowa State University

Towards a characterization of information automation systems on the flight deck

by

Rachel Feddersen Dudley

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Human Computer Interaction

Program of Study Committee:
Michael Dorneich, Co-Major Professor
Atul Kelkar, Co-Major Professor
Richard Stone

Iowa State University

Ames, Iowa

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

ABSTRACT

This thesis summarizes research to investigate the characteristics that define information automation systems used on aircraft flight decks and the significant impacts that these characteristics have on pilot performance. Major accomplishments of the work include the development of a set of characteristics that describe information automation systems on the flight deck and an experiment designed to study a subset of these characteristics. Information automation systems on the flight deck are responsible for the collection, processing, analysis, and presentation of data to the flightcrew. These systems pose human factors issues and challenges that must be considered by designers of these systems.

Based on a previously developed formal definition of information automation for aircraft flight deck systems, an analysis process was developed and conducted to reach a refined set of information automation *characteristics*. In this work, characteristics are defined as a set of properties or attributes that describe an information automation system's operation or behavior, which can be used to identify and assess potential human factors issues. Hypotheses were formed for a subset of the characteristics: Automation Visibility, Information Quality, and Display Complexity. An experimental investigation was developed to measure performance impacts related to these characteristics, which showed mixed results of expected and surprising findings, with many interactions. A set of recommendations were then developed based on the experimental observations.

Ensuring that the right information is presented to pilots at the right time and in the appropriate manner is the job of flight deck system designers. This work provides a foundation for developing recommendations and guidelines specific to information automation on the flight deck with the goal of improving the design and evaluation of information automation systems before they are implemented.

CHAPTER 1

INTRODUCTION

In order to safely and efficiently accommodate an increasing demand for air travel, the Next Generation Air Transportation System (NextGen) will utilize satellite-based navigation and interconnected database systems to guide and track air traffic more precisely than was previously feasible (FAA, 2013a). This system will integrate weather, traffic, terrain, and aircraft performance data to enhance safety while reducing delays, fuel requirements, and aircraft emissions. This transformation will result in increasing automation to take advantage of the likely increase in the amount of available information (Landry, 2009). Conveying the right information at the right time to the flightcrew and accepting input from them in a user-friendly manner is critical for safe operations.

Information automation systems collect, process, analyze, and present information to the flightcrew to support their task performance, decision making, and position awareness. Glass cockpit displays currently in use in many commercial air transport aircraft are examples of information automation systems. Their flexibility allows for any of the available information to be processed, analyzed, and presented to the flight crew whenever and however interface designers deem it appropriate. The primary goals of information automation systems are to promote situation awareness and assist in decision making tasks for the flightcrew; they are not intended to directly control the aircraft or its subsystems. Situation awareness and decision making assistance are specifically related to human information processing and cognition, while direct control of the aircraft and its subsystems is more heavily dependent on psychomotor skills and strategic mission planning. As such, there are likely unique human factors issues that must be considered when designing the interaction behavior of information automation systems.

This work was funded by the Federal Aviation Administration (FAA) Human Factors Division (ANG-C1) (contract #13-G-003). It was a conducted in close collaboration with the Human Centered Systems group at Honeywell Aerospace Advanced Technology in Minneapolis, MN, though under separate contract.

1.1   Problem Statement

Glass cockpit displays are very sophisticated information automation systems. As manufacturers develop these devices and their applications to incorporate and accommodate the NextGen directives, they must consider the implications of human factors issues in the design of the interactions and the presentation of information. While there is extensive literature on human factors issues related to aircraft automation in general, there is typically no distinction made regarding different types of automation (e.g., Tenney, Rogers, & Pew, 1998; Funk, Lyall, Wilson, Vint, Niemczyk, Suroteguh, & Owen, 1999), although there may be different human factors issues depending on the type of automation being considered. For example, Fadden (1990) and Billings (1991) introduced the concepts of control automation, management automation, and information automation. The PARC/CAST Flight Deck Automation Working Group has recommended that a stronger definition of information automation is needed, as well as definitions of terms related to it (FAA, 2013b). To address these recommendations, a more precise definition and characterization of information automation systems is needed in order to distinguish them from control and management automation systems.

Additionally, a thorough understanding of the human factors issues associated with the characteristics of information automation systems is also needed in order to enhance human performance and pilot interactions with these systems. This understanding will help prioritize those characteristics that have the greatest impact on pilot performance and will help guide the design decisions regarding what, when, and how data is presented. Given that tasks and priorities change throughout a mission, the impacts of the characteristics are also likely to change, depending on the context and situation the flightcrew is experiencing. For example, during approach, impacts such as workload and time pressure are much more important than when in cruise. An all-encompassing set of characteristics and a framework within which the characteristics can be described will help designers of information automation systems appropriately accommodate the human factors impacts associated with these changing environments.

Identification of the characteristics that describe information automation systems will help guide the metrics that can be used by design teams to work toward tangible, well-defined system specifications.

1.2   Research Objectives

The objectives of the work performed were to:

1) Generate a set of characteristics that describes flight deck information automation systems;

2) Generate and test hypotheses about the human factors impacts of a subset of key characteristics; and

3) Formulate design recommendations for information automation systems.

The focus of this work is in the domain of commercial transport flight decks, but in developing the hypotheses for evaluating the human factors impacts of information automation systems, previous work in other domains was also explored for broader perspective. The amount of information available to pilots is increasing and the importance of each piece of information can vary over the course of a flight, making information automation in this domain a particularly challenging area of research.

1.3   Thesis Organization

The structure of the thesis is as follows: this chapter introduces the motivation for the research, the problem statement, and the objectives of the research. Chapter 2 provides the relevant background information regarding automation and its human factors impacts, the types of automation encountered in the aviation domain, and a formal definition and framework for information automation in the aviation domain.

Chapter 3 presents the process followed in defining the characteristics of information automation in the aviation domain. It begins with a brief description of the initial, heuristic means to define the characteristics of information automation systems. Next, an analytical method to refine the characteristics is described. The chapter closes with the final proposed set of information automation system characteristics in the domain of commercial transport flight decks.

Chapter 4 focuses on three particularly important characteristics identified in the previous chapter: information quality, automation visibility, and display complexity. As each of these characteristics occupy their own areas of research, this chapter provides background information on each of these topics to describe how these characteristics are assessed in other

domains. The adaptation of these characteristics to the domain of interest in this research is also briefly introduced.

In Chapter 5, a description of the experimental method used to evaluate the three characteristics of interest is presented. Participants were given a decision making task, using an automated aid developed specifically for this research. Details are provided regarding the decision aid, the testing environment, the manipulation of the independent variables, and the measurement of the dependent variables.

Chapter 6 presents the results of the experiments and preliminary interpretations.

Next, Chapter 7 provides a discussion of the results. The human factors issues impacted by each of the characteristics tested are discussed at length and recommendations for design based on the experimental results are provided.

Finally, Chapter 8 summarizes the research and recommends future work.

CHAPTER 2

RELATED WORK

This chapter presents an overview of earlier work to describe automation systems in general and the potential human factors issues that can result from implementing such systems. Next, previous work to describe the various categories of automation specific to the aviation domain are discussed, leading to a formal definition of information automation on the flight deck as well as a framework to define flight deck information automation.

2.1   Automation

Automation has been defined as "…a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator" (Parasuraman, Sheridan, & Wickens, 2000, p. 287). Increasing computational capability and continuing technological innovations has led to sophisticated automation systems across many complex domains such as aviation, medicine, nuclear power, and manufacturing, just to name a few. Some of the obvious benefits of automation are increased reliability, efficiency, and throughput capability, while at the same time reducing both the physical and mental workload for their human operators.

There have been, however, some unanticipated consequences because of automation that have led to catastrophic events when the interactions between the human and the machine do not go as planned, or when the automation fails and the human operator is unable to intervene and recover from the failure. Two relatively recent aviation-related examples of such events are the Asiana Airlines flight 214 and Air France flight 447 accidents. A significant contributing factor in the Asiana 214 accident was confusion by the flightcrew regarding the autothrottle system during the approach to landing – "the flightcrew believed the autothrottle system would maintain the command speed" when, in fact, the autothrottle had been inadvertently deactivated by earlier actions by the pilot (Asiana Airlines, 2014, p. 32). In the Air France flight 447 accident, the airplane encountered high altitude ice crystals upon entering a line of thunderstorms at cruise altitude, resulting in blockage of all three pitot tubes and subsequent loss of airspeed measurement. The loss of airspeed measurement resulted in the autopilot disconnecting and the flight controls reverting from "normal law" to

"alternate law". Due to the unlikely event of either of these conditions, let alone both at once, the pilots had no experience with the handling qualities of the airplane under these circumstances (et d'Analyses, 2009; Palmer, 2013). These tragic events highlight the need for better understanding and accommodation of the human operator within complex systems.

An initial consideration in the design of automation systems is in the allocation of duties between humans and machines. Early work in this area includes the Fitts' list, also known by the acronym MABA-MABA ("Men are better at, Machines are better at"), which consists of six statements that point out those areas where humans perform better than machines and five statements indicating tasks that machines perform better than humans (Fitts, 1951). While this list is based on the technology available at the time, it continues to be cited today and has been argued as still being relevant to automation design (de Winter & Dodou, 2011).

Function allocation and Fitts' list then led to the concept of Levels of Automation, suggested by Sheridan and Verplank (1978) and later formalized by Sheridan (1980) and Wickens, Mavor, and McGee (1997). At the low extreme (level 1) of the ten-point Level of Automation scale, the human performs all tasks continuously with no computer assistance. At the other extreme (level 10), the computer makes all decisions and carries out the execution of those decisions with no human input. Between the extremes are different degrees of participation in a particular task by the human and the automation. Table 1 shows the Level of Automation scale as presented by Wickens, Mavor, Parasuraman, and McGee (1998).

Table 1. Scale of Levels of Automation of decision and control action (Wickens et al., 1998).

HIGH	10.	The computer decides everything and acts autonomously, ignoring the human.
	9.	informs the human only if it, the computer, decides to
	8.	informs the human only if asked, or
	7.	executes automatically, then necessarily informs the human, and
	6.	allows the human a restricted time to veto before automatic execution, or
	5.	executes that suggestion if the human approves, or
	4.	suggests one alternative, and
	3.	narrows the selection down to a few, or
	2.	The computer offers a complete set of decision/action alternatives, or
LOW	1.	The computer offers no assistance: the human must take all decisions and actions.

Parasuraman et al. (2000) refined the Level of Automation scale to include a second dimension to represent specific information processing stages:

1) Information acquisition;

2) Information analysis;

3) Decision and action selection; and

4) Action implementation.

In this model, levels of automation between the human and the system can be individually assigned at each of these four stages. Parasuraman et al. (2000) succinctly refer to these stages of automation as *acquisition*, *analysis*, *decision*, and *action* automation. Furthermore, the first two stages of acquisition and analysis automation are jointly referred to as *information* automation by the authors. The primary objective of information automation systems in this context is to augment the operator's perception and cognition.

Examples of *information acquisition* automation include mechanically manipulating sensors, organizing incoming information, prioritizing information, or filtering incoming information based on some criterion (Parasaurman et al., 2000). Examples of *information analysis* automation include (Bass & Pritchett, 2008): converting raw sensor data into an easier-to-understand form; comparing current sensor data to stored data or modeled predictions to assess performance or detect abnormal conditions; detect, predict, or highlight trends, patterns or conditions; or aggregating multiple information sources into a unified assessment.

Parasuraman et al. (2000) go on to define *decision* automation as consisting of various levels of assistance provided to the operator for making decisions. An example of *decision* automation include systems that have preprogrammed conditional logic rules built in that prescribe specific decision choices based on the existence of a particular set of conditions. Finally, an example of *action* automation is a system that would carry out or execute a selected response. These systems would typically replace the operator's hand or voice in the actual implementation of a decision.

2.2   Human Factors Impacts of Automation

Billings (1991) and Norman (1993) argued that the design of automation systems should be centered on the human operator, rather than pushing the human operator to the

periphery and forcing them to adapt to the automation. Wickens (1994) pointed out that a potential result of poor automation implementation is human operators being "out-of-the-loop" with what the system is doing, which compromises situation awareness, increases complacency, and may lead to degradation of domain-relevant cognitive reasoning skills. Therefore, automation strategies must be carefully designed *for* the operator, with the goal of keeping operators appropriately engaged in their tasks and goals.

While this philosophy has been widely agreed upon, its implementation has progressed rather slowly. Sheridan (2001) points to the difficulty of creating predictive models of human behavior over those of physical systems as a cause for this slow progression. Additionally, economic factors and rapidly emerging technology have continued to be the driving forces behind automation systems, resulting in a shift of human roles and responsibilities to essentially that of monitor, error handler, and automation manager (Sarter & Woods, 1997; Kaber, Wright, Prinzel, & Clamann, 2005), roles for which it is known that humans are not well suited (Wiener & Curry, 1980; Parasuraman, 1987). In these new roles, if an operator is not informed of what the system is doing or such indications are missed, then the operator may be surprised and perceive the system as behaving illogically. "Automation surprises" (Sarter, Woods, & Billings, 1997) occur when the system fails to take an expected action, or the automation carries out an action not explicitly commanded nor expected by the operator. This can lead to operators wondering what the system is doing and why, or what it will do next (Wiener, 1989).

From the performance aspect, the end result of automation surprise is typically delayed response or completely missing the opportunity to provide corrective action. Sarter and Woods (2000) conducted an experimental study with Airbus A-320 pilots in a full-flight simulator and demonstrated that pilots had more frequent instances of delayed interventions and errors of omission when interacting with systems with higher levels of autonomy and authority. In their summary, the authors point out that the difference between responses being delayed vs. completely missed was primarily dependent on whether effective feedback was provided. The authors also note that mode awareness problems can be addressed by making automated systems more observable. In particular, when the automation carries out an uncommanded action or transition, the system should actively alert the pilot to the situation.

The design of automated decision aids should include considerations regarding how much information is made available to the operator about the rationale, criteria, uncertainty, and determining factors used in forming the aid's judgments and its actions (Bass & Pritchett, 2008). The uncertainty considered by the automation, and how that uncertainty is communicated to the human, also impact operator decision making (Andre & Cutler, 1998) and performance (Bisantz, Marsiglio, & Munch, 2005). In addition, the human-automation interaction is complicated by a feedback loop between the automation's judgments and the human's information seeking, cue utilization, and judgment policy (Bass & Pritchett, 2008). If the algorithms used by the automation are highly complex and are dissimilar from the human's strategies or not understood by the human, the automation's outputs may be ignored (Adelman, Christian, Gualtieri, & Johnson, 1998; Kirlik, 1993). On the other hand, overly simplistic strategies may be disregarded as nuisances (Seagull & Sanderson, 2001). The type and level of information about automation reasoning and behavior has a strong effect on the human's trust, and may result in under or over-reliance on automation (Lee & See, 2004; Seong & Bisantz, 2002).

The type of automation may also lead to differing impacts in terms of human adaptability in using information automation. For instance, Kaber et al. (2005) found that for adaptive automation, humans were better able to adapt to changes in information analysis and action automation rather than for more cognitively intense information analysis and information decision automation.

2.3   Automation in the Aviation Domain

When considering the automation found on the aviation flight deck, Fadden (1990) provided an initial distinction of aviation automation into two main categories: *information automation*, which involves the management and presentation of context-relevant information to the flightcrew, and *control automation*, which addresses the automation of those devices that directly impact the aerodynamics of the aircraft. Billings (1997) introduced a third category of automation called *management automation*, which deals with the efficient completion of a mission. While control automation is clearly distinct from information and management automation, further details to distinguish these latter two are necessary.

According to Billings (1997, p. 70), information automation is "devoted to the management and presentation of relevant information to flight crew members". Examples of information automation systems include the following (1997, p. 88-105):

- Attitude and flight path displays

- Navigation displays

- Power displays

- Alerting and warning systems

- Communication automation

By contrast, management automation corresponds to the "strategic, rather than tactical, control of an operation" (Billings, 1997, p. 70). Management automation includes those functions allocated to the Flight Management System (FMS) for mission optimization (Billings, 1997, p. 109-110):

- Navigation: determination of position, velocity, and wind; management of navigation data sources.

- Aircraft system performance: trajectory determination, definition of guidance and control targets, flight path predictions; time and fuel at destination.

- Guidance: error determination, steering, and control command generation.

- Electronic instrument system: computation of map and situation data for display.

- Control-display unit: processing of keystrokes, flight plan construction, and presentation of performance and flight plan data.

- Input/output: processing of received and transmitted data.

- Built-in test: system monitoring, self-testing, and record keeping.

- Operating system: executive control of the operational program, memory management, and stored routines.

Between Billings' two lists of functions, there is an emphasis on the display, or presentation, of information in information automation that is not as prevalent (although still present) in management automation. Billings also notes the differences between information and management automation with respect to the types of computations performed by each system. Whereas management automation is heavily focused on strategic optimization tasks,

information automation has a focus on integration of available data sources into displays that aid broad functions such as task performance, decision making, and position awareness.

To summarize, information automation differs from control automation in that information automation has no direct impact on the aerodynamics of the aircraft, whereas control automation does have direct and immediate impact. The unique distinction of management automation is that it is focused on the strategic, rather than tactical, control of the aircraft in order to optimize performance over the course of the entire mission. Information automation systems are therefore explicitly used for the presentation of data in a timely manner and at the appropriate levels of abstraction for the task at hand.

2.4 Flight Deck Information Automation Definition and Framework

In order to focus the effort of characterizing information automation systems, a more formal and comprehensive definition of information automation was needed than what had previously been defined by Billings. Keeping the distinctions from the previous section in mind, a formal definition of information automation on the flight deck was developed: *Information automation encompasses all aspects of data collection (e.g., from sensors, databases, or human input), processing (filtering, prediction from models, varying levels of abstraction, etc.), and presentation to the human operator(s) through any appropriate modality (e.g., visual, auditory, and tactile).*

The three different categories of aviation automation specified by Billings (1997) and the four information processing stages specified by Parasuraman et al. (2000) led to the framework developed by Rogers, Whitlow, Letsu-Dake, Ott, and Dorneich (2013), which is shown in Figure 1. The horizontal dimension of the framework shows "What is controlled or acted upon?" The columns represent parameters similar to the aviation automation categories identified by Billings (1997) and reflect what the automation is controlling: the aircraft, the mission, or information. The leftmost column lists the "Information Processing Steps," and shows what stage of information processing is being performed by the automation. The steps were defined using the terminology from Boyd's Observe, Orient, Decide, and Act model (the OODA loop; Boyd, 1987). The rows of the table can be further identified as the four types of automation specified by Parasuraman et al. (2000): acquisition automaton (Observe), analysis automation (Orient), decision automation (Decide), and action automation (Act).

Different human factors issues are possible depending on the stage of information processing being performed.

What is Controlled or acted upon? / Information Processing Step	Aircraft Anything that affects aircraft performance	Mission Systems that do not directly impact aircraft performance (e.g., navigation, surveillance, communication)	Information Display systems that process and present information to the flightcrew
Observe detect, measure	Information automation linked to control automation	Information automation linked to management automation	Information automation
Orient interpret data, determine relevance, relate to situation and goals	Information automation linked to control automation	Information automation linked to management automation	Information automation
Decide evaluate options, formulate procedure	Control automation	Management automation	Information automation
Act execute	Control automation	Management automation	Information automation

☐ = Information automation
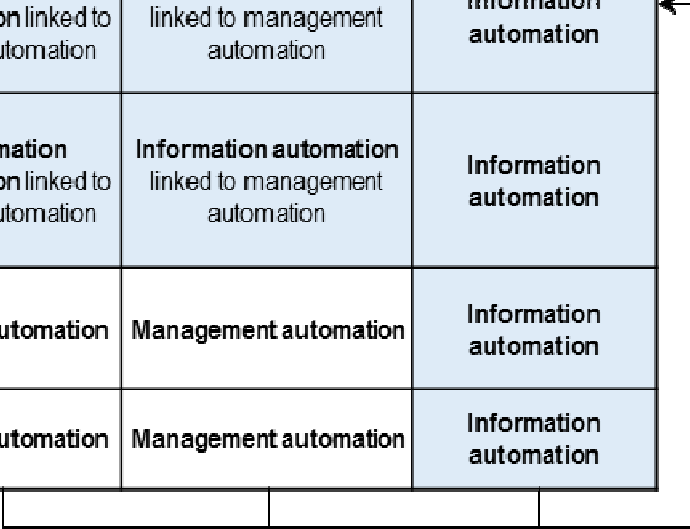
Feedback loops with processed information automation (this is considered information automation also)

Figure 1. Framework to distinguish information automation from control and management automation (from Rogers et al., 2013).

In the framework presented here, the definition of information automation is expanded to include not only the first two stages of processing, but also the final two stages if what is being controlled is information itself. For instance, information automation in the Orient/Information cell might provide judgments to a human operator (Bass & Pritchett, 2008), whereas information automation in the Orient/Aircraft cell might provide input into a hazard mitigation system that might affect the control of the automation. Both are considered information automation (specifically information analysis automation). Conversely, decision automation may or may not be classified as information automation. Automation in the Decide/Information cell that evaluates display options to decide the best way to convey information to the pilot would be information automation. Automation in the Decide/Aircraft

cell that decides on an evasive maneuver for the pilot would be considered control automation.

More specifically, the framework can be used to define areas considered to be information automation:

1) Early information processing stages (observe, orient) linked to control and management automation;

2) All information processing stages for automation where information is the primary commodity being controlled, processed, and presented; and

3) Feedback loops which present information on statuses and states for control and systems automation (while these loops might not be considered information automation per se, many similar human factors issues likely apply).

CHAPTER 3

CHARACTERIZATION OF INFORMATION AUTOMATION

This chapter describes the steps taken to develop a set of characteristics to describe information automation specific to aircraft flight deck systems. An initial brainstorming and categorization of information automation characteristics by Honeywell researchers is discussed first (Rogers et al., 2013). Their efforts laid the groundwork for the systematic analysis and refinement procedure performed as part of this research, which is then described. The goal of the characterization work was to establish a set of characteristics that would fully describe information automation systems on the flight deck without having any overlap in the characteristics; that is, each characteristic could be considered to be orthogonal to one another. Establishing such a set of characteristics could then allow for the development of metrics that could be used to objectively evaluate and compare different information automation system designs from a performance and usability perspective. For the purposes of this effort, a "characteristic" of an automated system was defined as an attribute, feature, or property which describes a system's operation or behavior.

3.1 Method

3.1.1 Initial Characteristics Generation

In addition to Honeywell's preliminary efforts in developing a framework for describing information automation systems on the flight deck, they also conducted several activities to identify characteristics of these systems that could lead to potential human factors issues. These activities included brainstorming meetings, pilot interviews, meetings with stakeholders and other human factors experts, a review of features of existing products, and a review of existing FAA design guidelines and recommendations. Multiple perspectives were considered:

- Products (e.g., Electronic Flight Bag applications)
- High level flight deck functions (e.g., aviate, navigate, and communicate)
- Flightcrew functions (e.g., communication with Air Traffic Control and Airline Operation Centers)

- Human error taxonomies (e.g., Threat & Error Management; Helmreich & Musson, 2000)

- Operational environment (NextGen; FAA, 2013a)

- Human information processing model (e.g., observe, orient, decide, act)

- Automation human factors (e.g., Billings, 1991; Parasuraman et al., 2000; Lee & See, 2004)

- Adaptive automation (e.g., Kaber et al., 2005; Feigh, Dorneich, & Hayes, 2012)

- Situation awareness (e.g., Endsley, 2000)

- User experience level (e.g., Rasmussen, 1983)

- FAA regulatory and guidance materials (e.g., Code of Federal Regulations, Advisory Circulars, and policy statements)

- Flight deck automation (e.g., Landry, 2009)

From these sources, the Honeywell research team generated an initial list of 130 features and attributes of information automation systems. Although using a multitude of perspectives created redundancy in feature identification, this redundancy was accepted in exchange for a more exhaustive analysis with a low probability of missing potential issues. The affinity diagramming process (Beyer & Holtzblatt, 1997) was used to organize the initial list of features and attributes into a hierarchy revealing common issues and themes. The affinity was built bottom up by collaboratively organizing related items, until all items were placed in groups. Categories for the groups were not predefined; rather they emerged from the contents of each group. The resulting list of candidate characteristics were then reduced to ten.

3.1.2 Characteristics Refinement

Refining the characteristics began by generating specific definitions for the ten characteristics in order to further analyze their independence, or orthogonality, from one another. The characteristics and their definitions are shown in Table 2 below.

Table 2. List of candidate characteristics of information automation.

Information Automation Candidate Characteristic	Definition
Complexity	Level of connectivity with other flight deck functions. Number and levels of automation present. Number of control and display elements required to interact with the system. The level of difficultly to understand the functions/sub-functions and what their current and future behavior will be.
Functionality	The intended function and the type of functions and their implications for risks from a human factors perspective. Potential for inducing distractions or being used for unintended functions. Frequency of information automation system use.
Authority	Level of authority/autonomy the system has over decisions and actions, even if those decisions and actions are only at the level of what/how/when information is presented. Amount of compellingness or salience that induces compliance and thus has implicit authority.
Level of Integration	Number of other systems or components directly linked to the system that have data or processing dependencies. Number of other systems that need to be evaluated in terms of consistency of user interface elements (colors, symbology, formats, etc.). Pilot procedures and operations that the system supports which require integration of new tasks with existing procedures.
Opacity	Ability for pilots to understand the system's behavior, how it is generating the outputs, and what sources it is using for input. Availability to verify its outputs. Ability to predict what it will do next.
User Interaction Requirements	Number and type of interaction required by the flight crew to successfully utilize the information automation system. Amount of head down time and/or distraction from other tasks. Amount of time to access information, provide inputs, or to interpret outputs from the system.
Criticality	Level of importance of the function that is supported from a safety perspective. Potential consequences if the system "gets it wrong."
Adaptiveness	Dynamic behavior of the system – level of ability to adapt its output to the situation such that it might appear more consistent and less predictable to the user.
Accuracy	Reliability, timeliness, and accuracy of the output. Ability to support the crew task. Potential to be misleading even if technically reliable.
Degradation Behavior	Failure modes of the system. Ability to easily identify and recover from failures. Existence of back up ways that pilots can achieve the same functions and outputs. Amount of risk of subtle and insidious failures and anomalies that might go undetected.

One limitation of the approach was that it was not capable of identifying whether any characteristics were missed. Generating the initial list by looking at the problem through different perspectives was an attempt to mitigate this issue. A second limitation was the possibility that some of the characteristics were redundant or captured similar human factors aspects of information automation. The analysis method described next was employed to address this limitation.

Rating Characteristics against Usability Principles

To ensure a level of independence between each of the characteristics defined in the previous section, a rating and correlation analysis was performed. This procedure was used in Dorneich, McGrath, Dudley, and Morris (2013) for an analysis of adaptive system characteristics. In that work, an initial set of 26 characteristics were reduced to a core set of seven, the independence of which the researchers had reasonable confidence in due to the analytic nature of the procedure. The method was adopted for this work to address similar concerns about the independence of the characteristics.

To evaluate their independence, or lack thereof, each of the characteristics were rated for the strength of their relation to each of the usability principles defined by Dix, Finlay, Abowd, and Beale (2004). These principles were chosen because they address three main categories of usability: *learnability*, *flexibility*, and *robustness*. Learnability affects the ease with which users can adapt their knowledge of current systems to a new interface. Flexibility deals with the various ways a user and system are able to exchange information. Robustness addresses a system's ability to support a user in assessing and achieving the user's goals. Within these three main usability categories are several principles, whose definitions are provided in Table 3. Together, these principles encompass all the important human-system interaction attributes of an interface and are therefore strong indicators for how readily an interface will be accepted and utilized by its users.

Table 3. Usability principles (from Dix et al., 2004, Ch. 7).

Principle	Definition
Learnability	
Predictability	Support for the user to determine the effect of future action based on past interaction history
Synthesizability	Support for the user to assess the effect of past operations on the current state
Familiarity	The extent to which a user's knowledge and experience in other real-world or computer-based domains can be applied when interacting with a new system
Generalizability	Support for the user to extend knowledge or specific interaction within and across applications to other similar situations
Consistency	Likeness in input-output behavior arising from similar situations or similar task objectives
Flexibility	
Dialog initiative	Allowing the user freedom from artificial constraints on the input dialog imposed by the system
Multi-threading	Ability of the system to support user interaction pertaining to more than one task at a time
Task migratability	The ability to pass control for the execution of a given task so that it becomes either internalized by the user or the system or shared between them
Substitutivity	Allowing equivalent values of input and output to be arbitrarily substituted for each other
Customizability	Modifiability of the user interface by the user or the system
Robustness	
Observability	Ability of the user to evaluate the internal state of the system from its perceivable representation
Recoverability	Ability of the user to take corrective action once an error has been recognized
Responsiveness	How the user perceives the rate of communication with the system
Task conformance	The degree to which the system services support all of the tasks the user wishes to perform and in the way that the user understands them

Each of the characteristics in Table 2 were rated by the strength of their relation to each of the usability principles in Table 3 for a 14x10 matrix, where the characteristics were the columns, and the usability principles were the rows. Three analysts individually rated each characteristic and usability principle combination (i.e., each cell of the matrix) on a scale of (0, 1, 3, 9). A nonlinear scale was used in order to emphasize the strength of the differences in the ratings. A rating of 9 represented a direct correlation where changes in the characteristic had a direct impact on the corresponding usability principle. A rating of 3 represented a strong relationship between the characteristic and usability principle, but with at least one other factor also affecting the usability. A rating of 1 was used to describe a weak relationship with several other factors affecting usability. Finally, a rating of 0 represented no relationship. For example, the complexity characteristic has future behavior of the system as part of its definition, so its relation to the predictability principle would be fairly strong.

The ratings by the three analysts were then compared and discrepant ratings reconciled through a series of meetings to discuss the rationale behind the individual ratings. It is important to note that the discrepant ratings were not averaged, rather consensus was reached through discussions in which example scenarios or anecdotes were considered. The reconciliation process allowed multiple perspectives to be considered that resulted in consensus between raters. In all cases, the analysts were able to reach consensus. As a measure of how consistent the participants were in their initial ratings, an inter-rater reliability analysis was also conducted.

Following the rating and reconciliation exercises, two analyses were performed on the data. The first was a measure of inter-rater reliability to determine how consistent the analysts were in their initial ratings. The second analysis was a Pearson's pairwise correlation analysis to assess the independence of each characteristic from one another. Linear independence of the characteristics' ratings along the 14 dimensions of the usability principles was estimated via Pearson's pairwise correlation analysis on each combination of characteristics. Each characteristic has 14 usability ratings. If one considers this a 1x14 "vector", then any two characteristics can be compared to see how similar their vectors are. A high correlation is an indication that two characteristics may be redundant. Similarly, high correlation of a characteristic to several others warrants further scrutiny to determine whether that characteristic should be modified, absorbed into one or more of the other characteristics,

or eliminated. Conversely, high correlation does not necessarily mean that a characteristic must be eliminated; rather, it signals areas that need further discussion. With $n(n-1)/2$ possible pairwise comparisons, even a moderate number of characteristics results in a significant number of comparisons, so the benefit of using this analytical method was to quickly identify those characteristics that needed further analysis from a human factors perspective without having to consider every combination.

3.1.3   Participants

Three human factors analysts participated in this analysis. The three analysts averaged 9.3 (range 6-15) years of aviation systems experience. In addition, one was a general aviation pilot.

3.1.4   Scope

It is important not to overstate the role that quantification (rating) of candidate characteristics played in this process. The ratings allowed a systematic comparison of candidate characteristics from a pilot perspective, and were used to guide the qualitative analysis of any correlations found. After human factors analysis, some correlated characteristics resulted in the characteristics being combined. However, there were also cases in which a quantitatively high correlation, after consideration and discussion, did not lead to a merging of characteristics. The goal of the quantitative (rating) exercise was to identify those combinations of candidate characteristics that warranted closer scrutiny; only the qualitative analysis determined the final disposition of the characteristics.

3.2   Results

The final, reconciled ratings between each of the characteristics and the usability principles is shown in Table 4. The sums and averages for each row and column are also presented in the table.

Table 4. Reconciled ratings of characteristics against usability principles.

Usability Principle \ System IA Characteristic	Complexity	Level of Integration	Opacity	Criticality	Accuracy	Functionality	Authority	User Interaction Requirements	Adaptiveness	Degradation Behavior	Sum	Average
Predictability	3	1	9	3	3	1	1	1	9	3	34	3.4
Synthesizability	3	1	9	3	3	3	1	1	3	9	36	3.6
Familiarity	3	3	3	0	1	3	1	3	1	3	21	2.1
Generalizability	3	9	3	0	1	3	1	3	3	3	29	2.9
Consistency	3	9	9	0	3	1	1	3	3	3	35	3.5
Dialog Initiative	1	1	1	1	0	3	1	3	1	0	12	1.2
Multi-threading	3	3	1	1	1	3	0	3	3	0	18	1.8
Task Migratability	1	3	3	3	0	3	9	3	9	1	35	3.5
Substitutivity	0	1	1	0	0	1	1	3	1	0	8	0.8
Customizability	1	1	1	1	1	3	3	9	9	1	30	3
Observability	9	1	9	3	1	3	3	1	3	9	42	4.2
Recoverability	3	3	3	3	3	9	3	3	3	9	42	4.2
Responsiveness	1	1	3	3	1	1	1	9	1	3	24	2.4
Task Conformance	3	3	3	9	9	9	1	1	9	3	50	5
Sum	37	40	58	30	27	46	27	46	58	47		
Average	2.64	2.86	4.14	2.14	1.93	3.29	1.93	3.29	4.14	3.36		

3.2.1 Inter-rater reliability

A measure of inter-rater reliability is helpful to understand the consistency of the participants' initial ratings. Inter-rater reliability was assessed by comparing each individual rating to the final reconciled rating and counting the number of "steps" between them. For instance, if a participant rated a cell as 9, a final rating of 3 would be one step away; a final rating of 1 would be two steps away; and a final rating of 0 would be three steps away. This method provides a conservative measure of the rate of agreement between participants. For

example, a set of ratings (3,3,1) may have been reconciled to a "1," so two participants were one step away from the final rating even though those two agreed with each other initially.

The final ratings matched 52.4% of the participants' initial ratings. Cumulatively, 93.3% of the participants' initial ratings were within 1 step of the final rating (see Figure 2), indicating that the reconciliation process to produce the final ratings started with a strong basis of agreement between analysts.
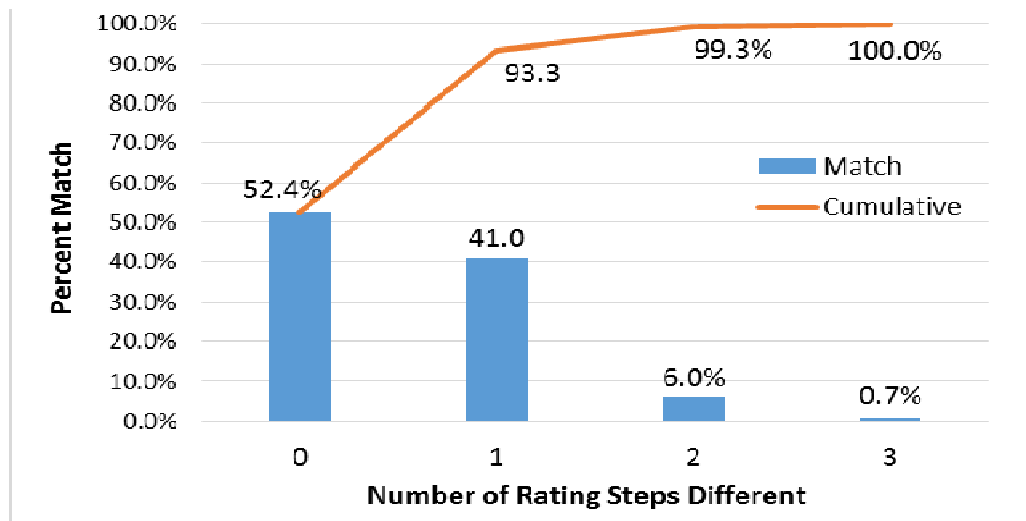
Figure 2. Participant rating match with final ratings.

3.2.2   Pearson's rank correlation

The Pearson's rank correlation for all combinations of characteristics is shown in Table 5. Correlations over 0.5 (typically considered strong) are in bold font. Correlations between 0.3 and 0.5 (typically considered weakly correlated) are in normal font. Correlations between 0.0 and 0.3 are in gray font. Since the relation between usability principle and characteristic is not symmetrical (i.e. a negative correlation does not imply any relation), any correlations below zero are blank. The final column is the "average correlation" for any one characteristic, which is the average of the correlations in the column above the diagonal cell element (denoted by "x") and the correlations in the row to the right of the diagonal element. As an example, Table 5 has the column/row combination for the Complexity characteristic outlined in bold.

Table 5. Pearson's rank correlation analysis, sorted from highest average correlation to lowest.

Characteristic	Criticality	Accuracy	Degradation Behavior	Functionality	Complexity	Opacity	Adaptiveness	Authority	Level of Integration	User Interaction Requirements	Average correlation
Criticality	x	**.80**	.29	**.64**	.19	.13	**.53**	.13			.27
Accuracy		x	.27	**.64**	.18	.25	.43		.12		.27
Degradation Behavior			x	.37	**.66**	**.62**		.02			.22
Functionality				x	.16		.25	.11	.01		.22
Complexity					x	**.61**			.07		.19
Opacity						x	.10		.12		.18
Adaptiveness							x	.47			.18
Authority								x		.06	.08
Level of Integration									x		.03
User Interaction Requirements										x	.01

3.3   Analysis

The correlations were analyzed in two ways. The first was to look at the single pairwise correlation. If two characteristics were highly correlated, then perhaps one could be eliminated as redundant, or the two could be combined into a single, more comprehensive characteristic. The data were also analyzed to study how much correlation there was between a single characteristic and all others as indicated by the "Average Correlation" in Table 5. In all cases correlations were used as indicators of necessary further discussion. A correlation by itself was not enough to eliminate a characteristic; a human factors basis for making a change to the characteristics was required.

Figure 3 illustrates the correlations between criticality, accuracy, and functionality. After discussion and analysis, it was decided that the relevant contextual aspects of task functionality and task criticality were supported by the quality of the information in the system (i.e. accuracy), and thus functionality and criticality were incorporated with accuracy

into the more broad characteristic of *Information Quality* (Wang & Strong, 1996), which includes the confidence that information meets intrinsic (including accuracy), contextual (including criticality and functionality), representational, and accessibility quality requirements.
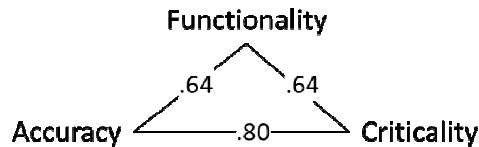
Figure 3. Correlations of accuracy, functionality, and criticality.

The next analysis considered the correlations between the characteristics of degradation behavior, opacity, and complexity (see Figure 4. Degradation behavior was considered a system characteristic, whereas opacity was considered a characteristic related to the interaction between the system and the human. Both were retained, although opacity was renamed *Automation Visibility* (e.g., Andre & Cutler, 1998; Bisantz et al., 2005). Complexity included both the functional complexity of information processing, as well as the level of complexity of information presentation. Complexity at the functional level was considered a system property, while complexity at the display level was considered more of a human-automation property. Rather than combining complexity with the other characteristics, it was split into two characteristics: *Functional Complexity* and *Display Complexity*. This is a good example of how the correlation method serves a triage method to identify areas where further analysis is needed. In this case, the analysis sparked by the high correlation results did not result in any candidate characteristics being eliminated.
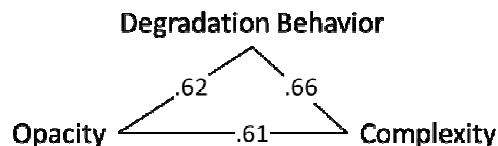
Figure 4. Correlations of opacity, degradation behavior, and complexity.

Authority was somewhat correlated (0.47) with adaptiveness. Authority is an emergent property of the function allocation, while adaptiveness of the system includes the function allocation. Therefore authority was eliminated as a redundant characteristic.

Finally, the analysis also revealed a natural grouping of characteristics between those associated with the automated system itself, and those associated with the interaction between the human and the system. Table 6 shows the final set of characteristics with their definitions grouped under these headings.

Table 6. Final set of information automation characteristics.

Information Automation Processing Characteristics	Definition
Functional Complexity	The complexity of the underlying processing and mode logic, and the understandability of functions and sub-functions.
Information Quality	The degree to which the information is fit for use; that is, the level of accuracy, completeness, timeliness, and so on, that can affect whether the information can be reliably used for the pilot task that it is intended to support.
Adaptiveness	How dynamic the system is – the degree to which it adapts its functionality, interaction, content, or task priorities to the situation such that it might appear less consistent and predictable to the user. The level of authority the system has over decisions and actions to adapt its behavior.
Level of integration	The number of other systems or components that are directly linked to the information automation system and have data or processing dependencies. The number of other systems that need to be evaluated in terms of consistency of user interface elements (colors, symbology, formats, etc.). The number of pilot procedures and operations that the information automation system supports which require integration of new tasks with existing procedures.
Degradation Behavior	The ways the system can fail or degrade. The degree to which the failure modes are easily detectable, easily reversible, and easily recoverable. The existence of back up ways the pilots can achieve the same functions and outputs. The level of risks of subtle and insidious failures and anomalies that might go undetected.
Human-Information Automation Interaction Characteristics	**Definition**
Display Complexity	The number of control and display elements that are used to interact with the system. The amount, variety, and organization of display elements that affect the pilot's ability to perceive, analyze, and act upon information.
Automation Visibility	The degree to which information is available to assist the user in understanding the system's behavior. The means by which the system provides information to allow the pilot to understand what sources of information the system uses as input, what reasoning it is using, and how it is generating the outputs. The methods provided to allow the flightcrew to verify its outputs and to predict what it will do next.
User Interaction Requirements	The amount and type of interaction that is required by the flightcrew to utilize the information automation system.
Compellingness	The level of attention and engagement that a system attracts (Wickens, Fadden, Merwin, & Ververs, 1998; Wickens & Alexander, 2009).

3.4   Summary

This chapter summarized the development of a formal definition and framework to describe information automation in the aviation domain. After an initial list of characteristics was developed, three research participants rated the relation of each characteristic to 14 usability principles. A Pearson's rank correlation analysis was then done in order to assess the independence of the characteristics. Where there was strong correlation among the characteristics, the analysis continued by considering the relationship(s) of the characteristics to one another from a systems and human factors perspective. Nine final characteristics were defined and grouped into system specific characteristics and human-system interaction characteristics. Some characteristics were eliminated, and (when appropriate), new characteristics were created to absorb or modify existing ones. The next chapter focuses on experimental evaluation of three of these characteristics: *Information Quality*, *Automation Visibility*, and *Display Complexity*.

CHAPTER 4

INFORMATION QUALITY, AUTOMATION VISIBILITY, AND COMPLEXITY

This chapter shifts the focus of the research to three characteristics of particular interest to FAA stakeholders: *Information Quality*, *Automation Visibility*, and *Complexity*. These three areas are recurring themes in the PARC/CAST Flight Deck Automation Working Group report (FAA, 2013b) and were deemed the highest priority for initial experimental study through a series of meetings between the researcher team, FAA program managers, and FAA technical sponsors. Some background information about these research areas as they relate to aviation and other complex domains is presented in order to inform a design of experiments. The details of the experiment and the results are reported and discussed in detail in Chapters 5 through 7.

4.1 Information Quality

Much of the previous research in what is referred to as *Information Quality* originated in database administration and management of information systems (Reeves & Bednar, 1994; Wang & Strong, 1996; Myers, Kappelman, & Prybutok, 1997; Pipino, Lee, & Wang, 2002; Stvilia, Gasser, Twidale, & Smith, 2007; Batini, Cappiello, Francalanci, & Maurino, 2009). In this domain, there are several factors that play an important part in the overall concept of information quality. For example, Wang and Strong (1996) identified four properties of high quality data: 1) intrinsically good, 2) contextually appropriate for the task, 3) clearly represented, and 4) accessible to the data consumer. They further identified 15 separate dimensions that fit within these four categories of information quality (see Table 7) in an effort to capture more comprehensively the *usefulness* of information as a product, or commodity, to the consumers who seek it.

Table 7. Categories and dimensions of information quality (Wang & Strong, 1996).

Intrinsically Good	Contextually Appropriate for the Task	Clearly Represented	Accessible to the Data Consumer
Believability Accuracy Objectivity Reputation	Value-added Relevancy Timeliness Completeness Appropriate amount of data	Interpretability Ease of understanding Representational consistency Concise representation	Accessibility Access security

English (2009) provided other components of information quality that covered similar underlying concerns, but in different groupings. These groupings were defined as: 1) quality of information product specification data, 2) quality of information content, 3) quality of information presentation, and 4) quality culture.

This categories and dimensions of information quality provided by Wang and Strong (1996) are directly relevant to pilots as consumers of the information provided by flight deck information automation systems and were therefore adopted for the experimental study. For example, automatic decluttering of a display based on the phase of flight and a predetermined set of criteria is a feature of adaptive automation that aims to provide pilots with only the most relevant and timely information for a given situation (Billings & Woods, 1994). A concern with this functionality, however, is whether the system is able to determine and provide all of the relevant information needed by the pilot for a given situation.

In the development of the experiments on information quality, several dimensions (e.g. accuracy, timeliness) could be manipulated. For example, introducing a delay in presented information would address the timeliness dimension. Furthermore, performance differences resulting from such manipulations may lead to recommendations for specific information quality dimensions.

4.2   Automation Visibility

Information automation visibility (sometimes also called "mode awareness" or "observability" in the literature – for example, Sarter and Woods, 1995; Woods, 1996; and Mosier et al., 2013) refers to the ability of an automation system to provide adequate feedback about its current state, what information is being used, and how the information is being processed (Endsley, 1996; Whitlow, Dorneich, Funk, & Miller, 2002). This characteristic may also be referred to as opacity (e.g., Andre & Cutler, 1998; Bisantz et al., 2005). In order for information automation to be visible, the feedback must provide a view into the automation's state and activities in a manner which can be properly interpreted by the operator (Woods, 1996) and allows the operator to predict its behavior (Scerbo, 1996).

In information automation systems that aid operators in decision-making tasks, good automation visibility would mean the system is effectively communicating what information it is using and how it is using that information to derive its recommendations. Many studies

have shown that providing meta-information and/or strategy information to operators improves task performance and error catching. For example, Seong and Bisantz (2008) found improvements in an air traffic identification task when the automation provided meta-information related to how the system applied a cue-weighting strategy to input data to come up with its judgments vs. providing its judgments without the underlying strategy. Other studies have shown decreased reaction times to alerts, along with improved responses to the alerts, when the automations' strategies were provided to operators when compared to the performance when strategies were not shown (e.g., Pritchett & Vándor, 2001; Sarter & Woods, 1992 and 1994a; and Skjerve & Skraaning, 2004). Building on these observations, Bass, Baumgart, and Shepley (2013) showed that judgment performance improvements could also be found in noisy environments when uncertainty information about the sensors that fed data to both the automation and the operators was provided as compared to the performance when operators were provided only the automation's judgment.

Highly automated systems that have low automation visibility may appear to the operator to be a completely autonomous agent, capable of its own independent actions. This is known as "perceived animacy" of the automated system (Sarter & Woods, 1994a) and on the flight deck can result in pilots having difficulty understanding system behavior when changing conditions cause a mode change that is not communicated effectively (Sarter & Woods, 1994b, 1995). For example, if a Flight Management System changes automation modes when a preprogrammed target altitude is reached and this change is poorly (or not at all) communicated, the pilot may perceive the system as acting on its own and wondering what its next actions will be. These situations of automation surprise are exacerbated as system complexity increases (Woods, 1996).

On the other hand, systems that have inappropriately high automation visibility may cause information overload (Deveans & Kewley, 2009; Degani, Barshi, & Shafto, 2013). Faced with an overwhelming amount of data, pilots may not be able to absorb the information presented to them and they run the risk of loss or reduction of situation awareness (Endsley, 1999, 2010; Wickens, 2002).

4.3   Complexity

From the previous work in defining the characteristics of information automation systems, two different complexity characteristics were identified: functional complexity and display complexity. The experimental study presented here focused on display complexity. This section gives an overview of complexity (in general) from the literature, followed by the dimensions and metrics adopted for measuring display complexity (specifically) of an information automation system.

The literature on complexity lacks a consensus on the definition of the term, although similar components in human-system interfaces have been identified (Cummings, Sasangohar, Thornburg, Xing, & D'Agostino, 2010). Three separate dimensions have been specifically recognized: *quantity* of basic information elements, *variety* of the elements, and the *relations* between the elements.

Boy (2008) interpreted perceived complexity as "complexity of an equipment or system in the flight deck as perceived by the pilot" (p. 8). He identified a broad range of issues related to artifact, user experience, task, organization, and situation complexities. Many of these components of perceived complexity relate closely to the usability principles given by Dix et al. (2004) used in refining the candidate information automation characteristics (see Table 3 in section 3.1.2). For example, artifact complexity is related to flexibility and task complexity includes consistency. Finally, Boy (2008) also points out that the main difficulty in measuring complexity is that it is related to expertise. Within this framework, then, complexity is a subjective measure that will likely vary from pilot to pilot.

However, an objective measure of complexity for information automation systems on the flight deck may be possible through the work by Xing (2007, 2008) in the domain of air traffic control systems. Xing first developed a framework for display complexity (Xing, 2007) and then developed a set of questionnaires to measure this type of complexity in air traffic control displays (2008). The wording in the questionnaires is sufficiently generic to be used in the evaluation of other types of displays as well. The framework consists of three basic factors: quantity, variety, and relation of information. Each of these factors is evaluated along three of the information processing stages: perception, cognition, and action. Additionally, the metrics are derived by associating the three complexity factors with the information processing stage (see Table 8).

Table 8. Display complexity dimensions and metrics as defined by Xing (2007).

	Perception	**Cognition**	**Action**
Quantity	No. of fixation groups	No. of functional units	Amount of action cost
Variety	No. of visual features	Dynamic complexity	Action depth
Relation	Degree of clutter	Relational complexity	No. of action goals

A multiple choice questionnaire was developed by Xing (2008) for quantitative evaluation of complexity of air traffic control displays. The questionnaire consists of a total of 13 questions: one question for each of the nine combinations of Table 8, followed by one question for each of the information processing stages (perception, cognition, and action), and a final question to address the overall display complexity. For each question, participants assigned one of the following four levels of complexity: 1) not complex, easy to use; 2) moderately complex but manageable; 3) complex and manageable only when not busy; 4) too complex to manage.

CHAPTER 5

EXPERIMENTAL METHOD

This chapter provides details of the experimental evaluation of varying levels of information quality, automation visibility, and display complexity on decision-making performance. The goal of the investigation is to gain insight into decision-making performance effects when these information automation system characteristics are manipulated. Understanding these effects will help establish design recommendations and guidelines for information automation systems.

5.1   Research Objectives

The goal of the experimental study was to show measurable differences in performance and other subjective assessment metrics when manipulating the information automation characteristics of information quality, automation visibility, and display modality.

5.2   Hypotheses

The evaluation is based on a premise that automation visibility will have an impact on the ability of pilots to detect problems resulting from poor information quality. The effect of display modality is also studied. Specifically, the hypotheses tested in the study were:

1. Increased information automation visibility will result in increased primary task performance, increased confidence in decisions, and increased trust in automation, but at a cost of higher workload.

2. Higher information quality will result in better primary task performance when compared to lower information quality.

3. Higher automation visibility will result in increased ability for pilots to compensate for poor information quality in the automation to maintain overall primary task performance (i.e., the difference in primary task performance between the low and high information quality conditions will be greater when automation visibility is low than when the visibility is high).

4. The graphical display will result in increased task performance, increased detection of information quality issues, lower complexity, and lower workload when compared with the text display.

5.3 Participants

Honeywell obtained Internal Review Board (IRB) approval for the study and recruited twelve airline transport pilots from a cross section of regional and major airlines to participate. All participants were right-handed males and one was color blind. Participants averaged 34.2 years of age (range: 24-56). Seven participants were First Officers and five were Captains. The average number of flight hours among the participants was 7000 (range: 2000 – 14,000). Seven of the pilots had no experience with electronic flight bags, four had some experience, and one used an electronic flight bags in his daily work. Participants rated their familiarity with glass cockpits as 4.9 of a 5-point scale (standard deviation 0.3). Also on a 5 point scale, participants rated their level of trust in automation at 3.83, with 1 being no trust and 5 being complete trust in automation (standard deviation 0.55). Finally, the pilots were asked about their level of authority in making decisions about diversions, with 1 being they had no authority and 5 being that they had complete authority. The average response among the pilots was 4.08 and the standard deviation was 1.08.

5.4 Experiment Task: Diversion Decision-Making

For this experiment, participants used an information automation system designed to aid in-flight diversion decision making. This section provides some background on diversion and a brief introduction to the task assigned to participants.

5.4.1 Background

Historically, diversion decisions have been a collaborative effort between the pilot and airline dispatchers. As more information becomes available on the flight deck with NextGen capabilities, the balance of responsibility for diversion decisions may shift more toward pilots. The primary goal of this task is to ensure that the plane is diverted safely. Secondary goals may include minimized downstream disruptions to airline operations. Experienced dispatchers know that diversion decisions have significant impact on downstream airline operations, including the schedules for aircraft, crew, maintenance, and

passengers (Dorneich, Whitlow, Miller, & Allen, 2004). In future operations, pilots may be expected to take a more active role in considering these aspects of diversion decisions.

For this study, participants were tasked with making a diversion decision based on the assumption that the diversion decision making tool had already considered all aspects of the flight related to safety, such as remaining fuel and runway lengths at the suggested diversion airports. As such, they were to focus on the consequences of the diversion options from an airline operations perspective. This was a modified way for pilots to consider the diversion decision-making task for two reasons: 1) they were to assume the safety requirements were met by the automation, and 2) they were not coordinating with airline dispatchers.

From the perspective of the dispatcher, diversion decisions consist of two parts: which of the in-flight aircraft are to be diverted, and to which airports they are diverted. These two decisions can have dramatic consequences in the disruption of an airline's four inter-linked schedules: aircraft, crew, maintenance, and passenger schedules.

There are other stakeholders in diversion decisions; however, the diversion decision is made by only the pilot and the dispatcher. In addition, there is very little time available to produce a diversion plan, which one dispatcher characterized as "0-10 minutes" (Dorneich et al., 2004). The relevant information about how a candidate plan will affect various schedules and their stakeholders is distributed across multiple systems and departments. Consequently, in current practice the decision is almost solely based on fuel limits and other aspects of aircraft safety. There are typically several different diversion plans possible that will maintain safe flight and landing profiles, but differ widely in their impact on airline operations, profits, crew and staff convenience, and customer satisfaction.

5.4.2   Policies

A set of company policy statements was established to represent the operational priorities of all stakeholders affected by diversion decisions. These policies are used to assess the overall "goodness" of a diversion plan. Each policy was associated with cost points operational for each statement that is violated by a particular plan. The policy statements are shown in Table 9. For example, diverting a flight with an unaccompanied minor costs 10 points, while delaying a flight greater than 15 minutes costs 8 points. The policy statements are adapted from a list of policy statements developed by Dorneich et al. (2004) after

conducting interviews with airline dispatchers as well as various stakeholders. The goal of selecting a diversion option is to minimize the total cost incurred by the selected option. The lower the cost, the better the plan.

Table 9. Policy statements and their corresponding cost values.

Policy Statement	Cost
Do not exceed crew duty limits	10
Do not divert a flight with an unaccompanied minor	10
Do not divert a flight with an arriving international passenger to an airport that does not have passport control	10
Do not divert passengers connecting to an international flight	8
Do not delay flights greater than 15 minutes	8
Do not divert to an airport that has its maximum capacity of aircraft	8
Do not cause crew to miss next flight assignment	5
Do not cause passengers to fail to reach destination	3

5.4.3 Diversion Aid

An information automation tool, the Diversion Aid, was created for the purposes of the study. The Diversion Aid integrates multiple information sources to provide participants with data on the current state of flight, aircraft, maintenance, crew, and passenger schedules. By capturing and showing the implications of diversion decisions to the participant, it was anticipated that s/he would be better able to integrate the goals and priorities of interested airline operations stakeholders into the decision making process.

5.4.4 Displays

The Diversion Aid presented the original scheduled flight plan, followed by its diversion plan recommendations to the participants in one of three ways (automation visibility options), depending on the experimental condition. Additionally, the Diversion Aid was presented in one of two display modes: *text* or *graphic*. The text and graphic displays

were designed such that the information content was identical between the two modes. Figures 5 and 6 show annotated descriptions of the text and graphic displays, respectively.

The automation visibility options were:

- Low Visibility: A single best option (see Figures 5 and 6)
- Medium Visibility: A ranked list of the top three options (see Figure 7)
- High Visibility: A ranked list of the top three options with the cost values shown (see Figure 8).
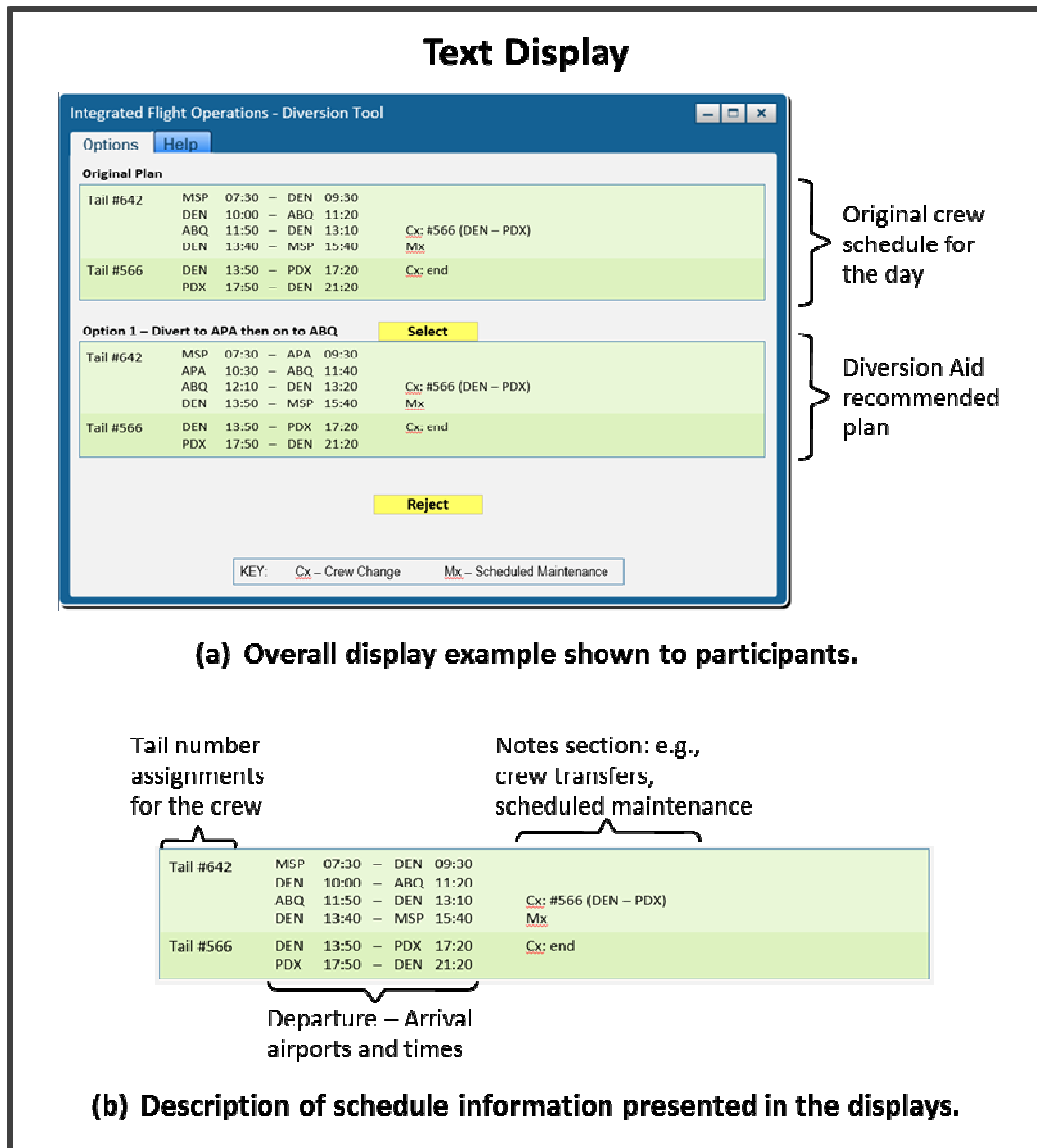
Figure 5. Annotated Diversion Aid presenting options with low automation visibility in text display form: (a) overall display example shown to participants; (b) description of schedule information presented in the displays.
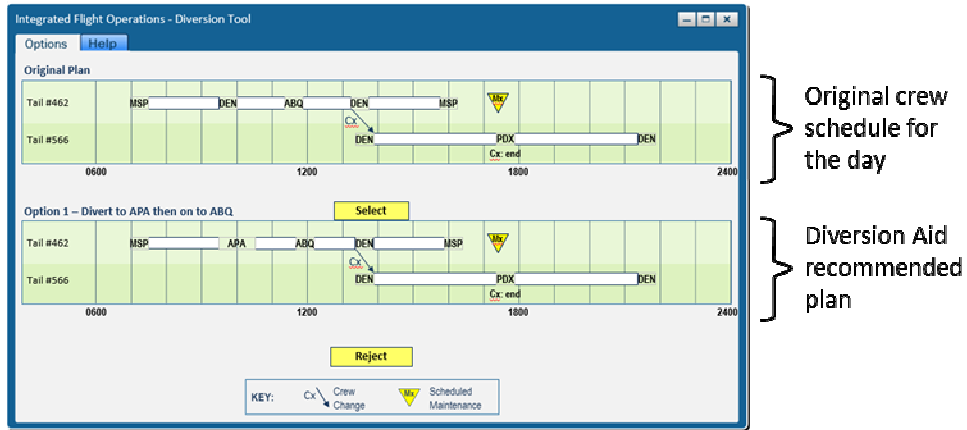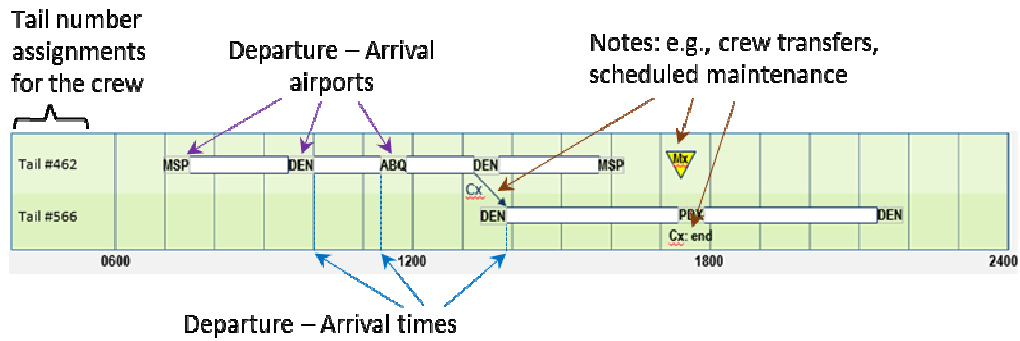
Figure 6. Annotated Diversion Aid presenting options with low automation visibility in graphic display form: (a) overall display example shown to participants; (b) description of schedule information presented in the displays.
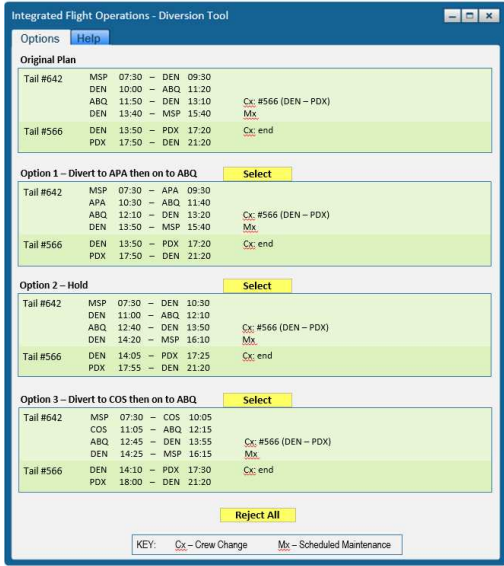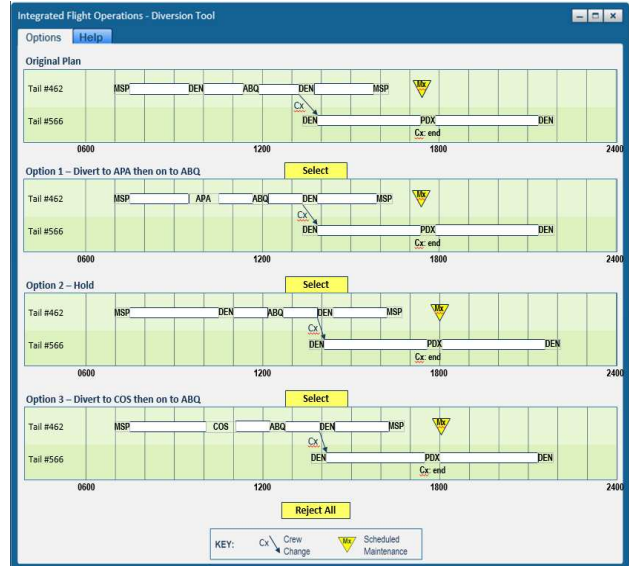
(text)　　　　　　　　　　　　　　**(graphic)**

Figure 7. Diversion Aid presenting options with medium automation visibility in text and graphic display forms.
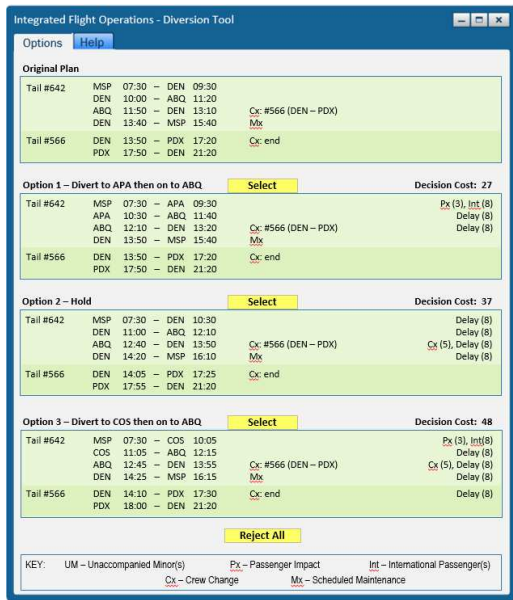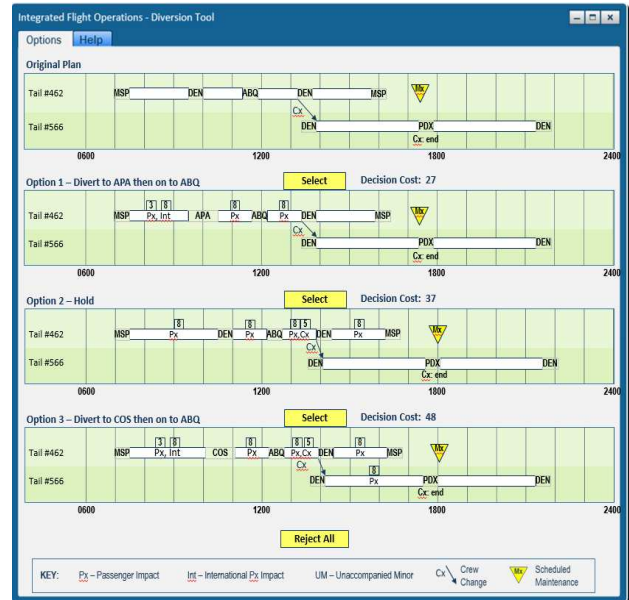
(text)　　　　　　　　　　　　　　**(graphic)**

Figure 8. Diversion Aid presenting options with high automation visibility in text and graphic display forms.

5.5   Tasks / Scenarios

Participants performed two tasks for the experiment: the primary task was to select or reject a diversion plan with the help of the Diversion Aid, and the secondary task was to report traffic as it appeared in an out-the-window display. In every trial, participants knew they would be diverted, but did not know when in the scenario they would be instructed to divert.

Participants acted as the Pilot Monitoring and performed six trials, each with a unique scenario that represented a typical crew schedule for one day, including up to one crew transfer to another aircraft (tail number). In an abbreviated pre-flight briefing, the confederate pilot reviewed the schedule for the day, weather, and a pre-planned diversion airport with the participant. These briefings contained both relevant and irrelevant information specific to the diversion task, in order to provide the information that might be needed to make a correct decision without explicitly stating that the information would be required. Participants were informed that the Diversion Aid may not always have the most current or correct information, in an attempt to appropriately calibrate trust. Participants were also told that the briefings had the most accurate and up-to-date information and they, as pilots, had the final authority in the diversion decision.

5.4.5   Primary Task

A flight simulation was presented to the participants to help provide a sense of realism to the trials. The simulation began approximately ten minutes from top of decent After 60 to 90 seconds, the need for a diversion was announced and the participant was asked to make a recommendation within five minutes. The participant then started the Diversion Aid, reviewed its recommended plan(s), and decided whether to accept one of the plans or to reject its recommendation(s) if he felt he could devise a better plan. The participant did not need to create a different plan. A help menu was available that displayed the set of policies (see Table 9).

5.4.6   Secondary Task

A secondary task of reporting traffic in a simulated out the window view was also assigned in order to increase workload during the diversion selection task. Traffic appeared

out the window (see Figure 9) in random locations at random times and did not move. Every five seconds, the probability of traffic being displayed was 60% (a set point determined during pre-experimental dry runs). If traffic did appear, it remained in the view until participants reported it by pressing a button (usually held in their non-dominant hand). If they failed to press the button after five seconds, the target would disappear. Participants were instructed to press the button as soon as they saw the traffic. While reviewing the Diversion Aid's recommendations, the participant continued reporting out-the-window traffic.
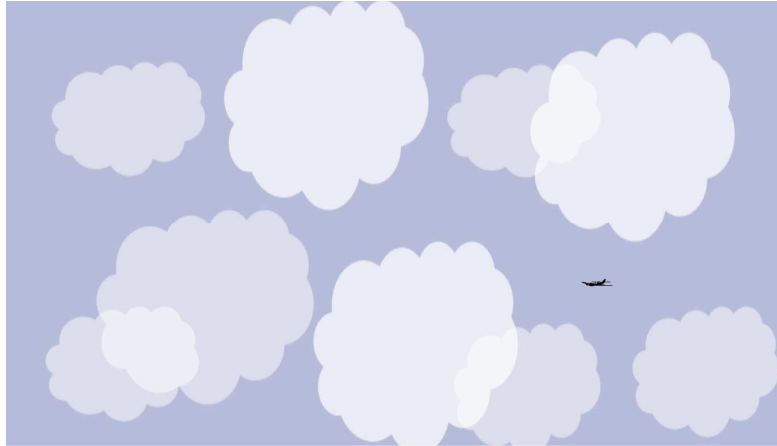
Figure 9. Simulated out the window view with traffic.

5.6 Independent Variables

In addition to the two Display Modes and the three Automation Visibility levels, Information Quality was also an independent variable in the experiment. The independent variables are summarized in Table 10.

Table 10. Independent variables for the experimental study.

Independent Variable	Levels	Description
Information Quality	Low	Some relevant information was not included in the calculation of total diversion decision cost
	High	All relevant information was included in the calculation of total diversion decision cost
Automation Visibility	Low	Best option only
	Medium	Rank-ordered list of the three best options
	High	Rank-ordered list with the costs shown
Display Mode	Text	Plan information was displayed in text form
	Graphic	Plan information was displayed in graphic form

5.7 Dependent Variables

Table 11 shows the dependent variables that were measured.

Table 11. Dependent variables and the metrics used to measure them.

Dependent Variable	Metric	Measurement / Unit	Frequency of Collection	Data Type
Decision Performance	Time to make a selection	seconds	once per trial	ratio
Decision Performance	Selection of best plan	yes/no	once per trial	binary ordinal
Workload	TLX measures: a) Mental Demand b) Physical Demand c) Temporal Demand d) Performance e) Effort f) Frustration	0 - 10	once per trial	subjective ordinal
Workload	Ratio of detected vs. all targets	%	2x per trial - before and after diversion selection task	continuous
Attention Allocation	Time spent on primary and secondary displays (app vs. out the window)	seconds	once per trial	ratio
Confidence	Survey question: Confidence in decision	1 - 5	once per trial	subjective ordinal
Automation Awareness	Survey question: Understanding of automation	1 - 5	once per trial	subjective ordinal
Trust	Survey question: Trust in automation	1 - 5	once per trial	subjective ordinal
Display Complexity	Survey questions on complexity: a) Perception - Quantity b) Cognition - Relation c) Overall Perceptual Complexity	1 - 4	once per trial	subjective ordinal

Decision Performance was measured via the time to make a diversion decision, and the correctness of the decision. Time to make a decision was the elapsed time from the start of the Diversion Aid until participants made their diversion plan selection. Participants were asked to select a diversion plan from the options presented by the Diversion Aid, or to reject all options if they felt that there was a better plan. Plan selection performance was scored as a 1 if the participant selected the best plan that resulted in the least cost according to the policy statements. If the best plan was not selected, the result was 0. In the high Information Quality

condition, the correct selection was always the top option on the display. In the low Information Quality condition, the automation was missing information that resulted in incorrect scoring of the options. Participants were briefed earlier and possessed this missing information. Thus in the low Information Quality conditions, the automation's highest ranked plan was not actually the best plan – participants were expected to recognize that a different plan was better once they included the missing information into to their assessment. They could also reject all the plans shown if they felt that the options shown were flawed. In the medium and high Automation Visibility conditions for the low Information Quality trials, this means that the actual best plan (correct selection) was listed below the automation's highest ranked plan; in the low Automation Visibility condition, there was only one option shown by the automation, so if the participant recognized that there was missing information, he could choose to reject the plan.

Workload while selecting a diversion plan was measured two ways. The first was subjective workload measured via the NASA-TLX questionnaire (Hart & Staveland, 1988), which assessed workload along six dimensions that were summed to arrive at a total workload value. The second measure was an objective measurement of workload based on performance of the secondary task of reporting traffic detected in an out-the-window view.

Attention allocation was estimated using head-tracking data to calculating the percentage of time the participant spent looking at the Diversion Aid while selecting a plan.

Confidence, Automation Awareness, Trust, and Display Complexity were measured via a post-trial questionnaire (see APPENDIX A). The Automation Awareness question asked participants to provide their level of understanding of how the Diversion Aid arrived at its recommendations. Three survey questions from Xing (2008) were used to assess participants' opinions on the complexity of the displays.

A post-experiment questionnaire was also administered (see APPENDIX B) to collect participants' qualitative responses regarding automation visibility, addressing what strategies they used to come up with their decision, what they liked, and what they would improve. Participants were asked to rate their relative preference between the three automation visibility levels by distributing a total 100 points for each of the following five attributes (with more points indicating higher preference):

- Clarity of information

- Completeness of information
- Ease of finding information
- Helpfulness in making a decision
- Preference

For example, a participant might allocate the following point for the "Completeness of information" attribute: 20 for low automation visibility, 30 for medium, and 50 for high.

5.8   Experimental Design

The experiment was designed as a 2 (Information Quality) x 3 (Automation Visibility) x 2 (Display Mode). Display mode was a between subjects variable, so participants saw either the text or graphic display mode, but not both. Information Quality and Automation Visibility were manipulated within subjects. Table 12 shows the treatment assignments for the participants. The odd numbered participants saw the text display while the even numbered participants saw the graphic display.

Table 12. Treatment assignments for each of the twelve participants. Odd numbered participants saw the text display while the even numbered participants saw the graphic display.

Scenario	Participant / Treatment	1	2	3	4	5	6	7	8	9	10	11	12
1	Visibility	High	High	Low	Low	Med	Med	Low	Low	High	High	Med	Med
	Info Quality	Low	Low	High	High	Low	Low	Low	Low	High	High	High	High
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic
2	Visibility	Low	Low	Med	Med	High	High	High	High	Med	Med	Low	Low
	Info Quality	High	High	Low	Low	High	High	Low	Low	High	High	Low	Low
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic
3	Visibility	Med	Med	High	High	Low	Low	Med	Med	Low	Low	High	High
	Info Quality	Low	Low	High	High	High	High	High	High	Low	Low	Low	Low
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic
4	Visibility	Low	Low	Med	Med	High	High	High	High	Med	Med	Low	Low
	Info Quality	Low	Low	High	High	Low	Low	High	High	Low	Low	High	High
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic
5	Visibility	Med	Med	High	High	Low	Low	Med	Med	Low	Low	High	High
	Info Quality	High	High	Low	Low	Low	Low	Low	Low	High	High	High	High
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic
6	Visibility	High	High	Low	Low	Med	Med	Low	Low	High	High	Med	Med
	Info Quality	High	High	Low	Low	Low	High	High	High	Low	Low	Low	Low
	Display	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic	Text	Graphic

The order of the scenarios were counterbalanced across participants to avoid learning effects that might otherwise have been associated with the scenario number.

5.9 Testing Environment

The evaluation was conducted in a low fidelity flight simulator in a Honeywell facility in Golden Valley, MN. Figure 10 shows the general layout of the simulator. Microsoft® Flight Simulator X (FSX) was used for the flight simulation. An InterSense® InertiaCube2 head tracker was used to measure percent time spent looking at different displays (Figure 11). A video camera was used to record participant interactions with the information automation applications.
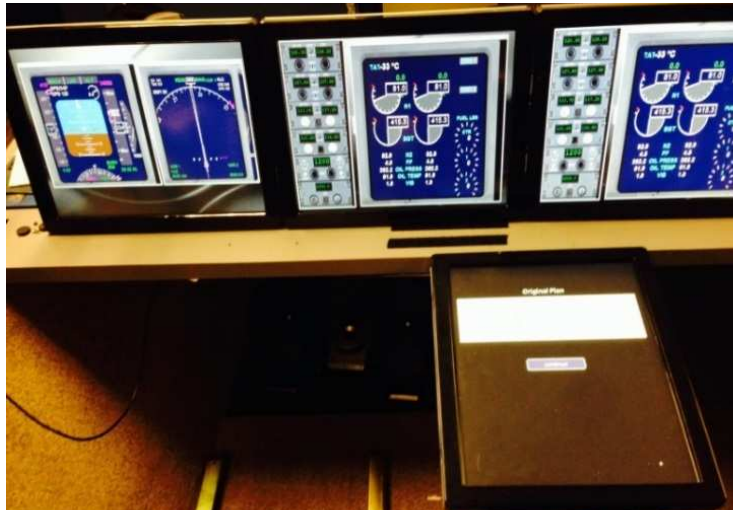

Figure 10. Low fidelity simulator layout.

Figure 11. Head tracker.

5.10 Procedure

Upon arrival, each participant was first given an initial briefing of the days' activities and a consent form to read and sign. A questionnaire was then given to gather demographics, piloting experience, use of electronic flight bags, and general attitudes toward automation.

Following the preliminary paperwork, participants were trained on the use of the Diversion Aid, the tasks they would be asked to perform, and the post-trial questionnaires they would complete. The first part of the training was conducted outside the simulator, with the experimenter reading a script while stepping through training slides that followed the script in order to provide all participants with the same information. The training included stop points at which the participants were asked to explain what information was being shown in the aid to ensure they had a reasonable understanding of how the aid calculated and presented its recommendations. The questionnaires that would be administered after each trial were also given to the participants so they could practice completing them before starting the trials.

The second part of the training was performed in the simulator by working through a training scenario with step-by-step instructions given. The conditions for the training scenario were set to high Information Quality and high Automation Visibility. Upon completion of the training scenario, if participants were able to make a diversion plan selection within a five minute time limit, it was determined that their performance was satisfactory and the actual trials began. If they required more than five minutes or still felt unsure about the task, the training scenario was repeated.

Participants completed a total of six different diversion scenarios in the simulator. After making each diversion decision, they filled out the NASA-TLX workload scale and post-trial questionnaire for each scenario. After all six scenarios were completed, participants filled out a post-experiment questionnaire and were provided a short debrief of the experiment.

5.4.7 Limitations and Assumptions

One limitation of the study was that the task required pilots to think about diversions in a completely different way than how they are used to handling them. Training, repeated

reminders, and practice runs were used to orient them to the Diversion Aid and all pilots were able to accomplish the tasks.

A second limitation of the study was the limited number of participants. Given the 2x3x2 experimental design and only 12 participants, statistical power of the experiment was anticipated to be low. The data was analyzed for statistical significance.

CHAPTER 6

RESULTS

Due to the limited number of participants and the relatively high number of manipulations (2x2x3), potential significance of the independent variables on each of the dependent variables was first investigated through a Least Squares analysis, with significance threshold alpha = 0.05 and marginal threshold alpha = 0.1. The initial statistical analysis of the Display Mode manipulation did not reveal significant results, so this independent variable dimension was collapsed in order to increase the power of the analysis.

Potential significance of the independent variables on each of the dependent variables was investigated through a repeated measures analysis of variance, where Information Quality and Information Automation Visibility were treated as the repeated measures. Results were considered significant for a threshold set to alpha = 0.05, and marginally significant for a threshold of alpha = 0.1. Table 13 shows a summary of the *p*-values, with the significant and marginally significant results in bold font and noted with (*****) and (**m**), respectively. Detailed results of the significant and marginally significant results are presented in the following subsections.

Table 13. Summary of p-values; (*) indicates a significant result, (m) indicates a marginally significant result.

Dependent Variable	Metric	Automation Visibility	Information Quality	Automation Visibility × Information Quality
Decision Performance	Selection of best plan	.734	**.00013 (*)**	.534
Decision Performance	Time to make a selection	**.042 (*)**	.118	.649
Subjective Workload	TLX Total Workload	.160	.463	.161
	Individual TLX measures:			
	a) Mental Demand	**.0511 (m)**	.300	.122
	b) Physical Demand	.146	.261	.214
	c) Temporal Demand	**.022 (*)**	.634	.889
	d) Performance	.913	.920	**.050 (*)**
	e) Effort	.119	.893	.649
	f) Frustration	.670	.529	.310

Dependent Variable	Metric	Automation Visibility	Information Quality	Automation Visibility × Information Quality
Objective Workload	Ratio of detected vs. all targets	.491	.072 (m)	.663
Attention Allocation	Time spent on primary and secondary displays (app vs. out the window)	.0005 (*)	.044 (*)	.838
Confidence	Survey question: Confidence in decision	.559	.067 (m)	.171
Automation Awareness	Survey question: Understanding of automation	.015 (*)	.067 (m)	.093m
Trust	Survey question: Trust in automation	.030 (*)	.031 (*)	.031 (*)

6.1 Decision Performance

6.1.1 Plan Selection

This measure was the percentage of trials that the participant chose the best plan. Information Quality was a significant manipulation for this measure ($F_{(1,11)} = 32.98$, $p < 0.00013$). Automation Visibility was not significant. Figure 12 shows the means and standard error of the correct selection percentage for the low and high Information Quality conditions.
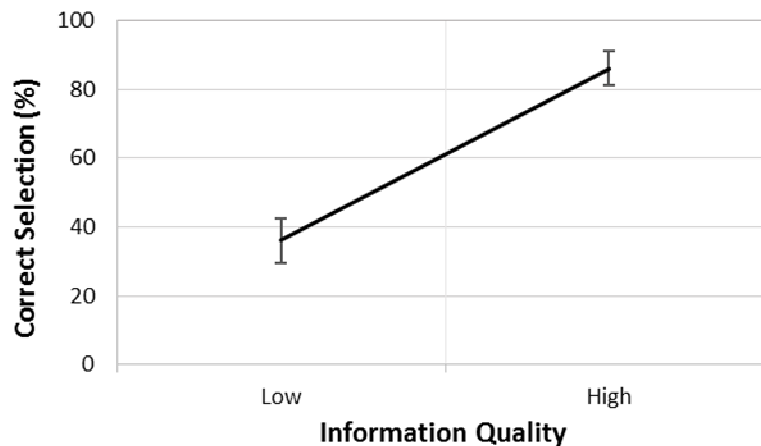

Figure 12. Mean and standard error for the correct selection percentage.

In each scenario, one of the plans considered by the automation was to hold and wait for a specified time given by air traffic control. All of the participants commented that they

were biased towards picking the hold plan if displayed as an option, despite being told to only consider the policy statements, because holding was much easier than diverting based on their operational experience. Diversions introduce new tasks, e.g., reviewing new approach charts, planning for a new and possibly unfamiliar airport, and making adjustments to their schedules. Participants' comments also suggested that they considered passenger impact in their decisions much more heavily than what the policy statements warranted. Thus, participants were prone to selecting the hold plan (if presented) over the top ranked option.

On average, participants correctly identified the best plan in 36% of the low Information Quality trials and 86% of the high Information Quality trials. Automation Visibility level was not significant, so this was not a driving factor in the participants' ability to catch the missing information. Overall, participants were not able to consistently detect missing information and incorporate that knowledge into their decisions.

6.1.2   Time to Make a Selection

The time to make a selection was the elapsed time from the start of the Diversion Aid until participants made their diversion plan selection. Automation Visibility was a significant factor ($F_{(2,22)} = 3.67$, $p < 0.042$) for this measure, with the low Automation Visibility condition being significantly faster ($t_{(22)} = 2.15$, $p < 0.043$) than the high Automation Visibility condition. Figure 13 shows the time to make a selection as a function of Automation Visibility.
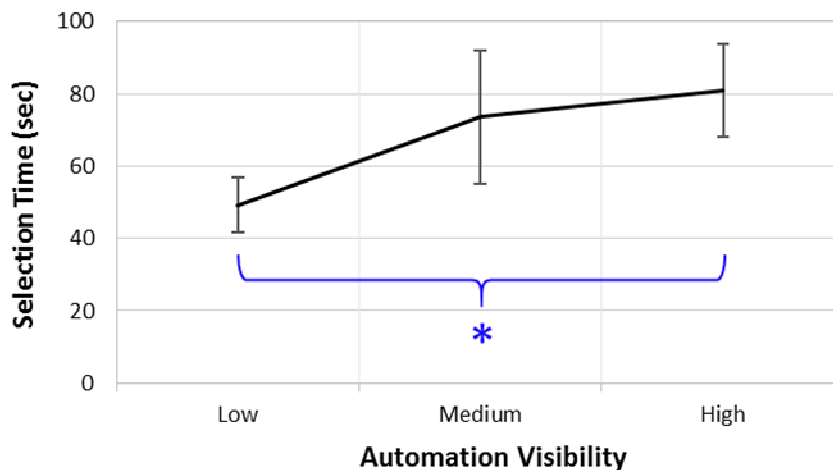

Figure 13. Mean and standard error for diversion plan selection time.

Plan selection times were shorter under the low Visibility level because there were fewer options to consider and less information to process and decipher. One participant commented that he preferred either the low or the high Automation Visibility level, as the medium Automation Visibility level (ranked options *without* cost) were too much work to interpret:

> "The single option was superior to rank ordered because a decision made
> without seeing the reason can just be a suggestion. Three suggestions without
> the reason behind add more workload. Having the decision cost allowed
> quicker decision making and a more informed decision."

6.2   Workload

Two workload measures were gathered in the experiments. The first was a subjective workload, measured via the NASA-TLX questionnaire, which assess workload along six dimensions that are summed to arrive at a total workload value. The second measure was an objective measurement of workload based on performance of the secondary task of reporting traffic detected in an out-the-window view.

6.2.1   Subjective Workload (NASA-TLX)

The total NASA TLX workload showed no significant results for any of the dependent variables, nor their interaction. There were, however, individual TLX measures that showed significant or marginally significant results.

Mental Demand: Automation Visibility had a marginally significant ($F_{(2,22)} = 3.41$, $p < 0.051$) impact on the mental demand results. Looking further into a paired *t*-test between the three Automation Visibility levels shows significantly lower ($t_{(22)} = 2.10$, $p < 0.047$) mental demand for the low Automation Visibility level vs. the medium level (see Figure 14). With only one plan presented by the automation, the only two choices would be to either accept the one provided or to reject it, resulting in less mental demand required compared to evaluating the costs of multiple plans.
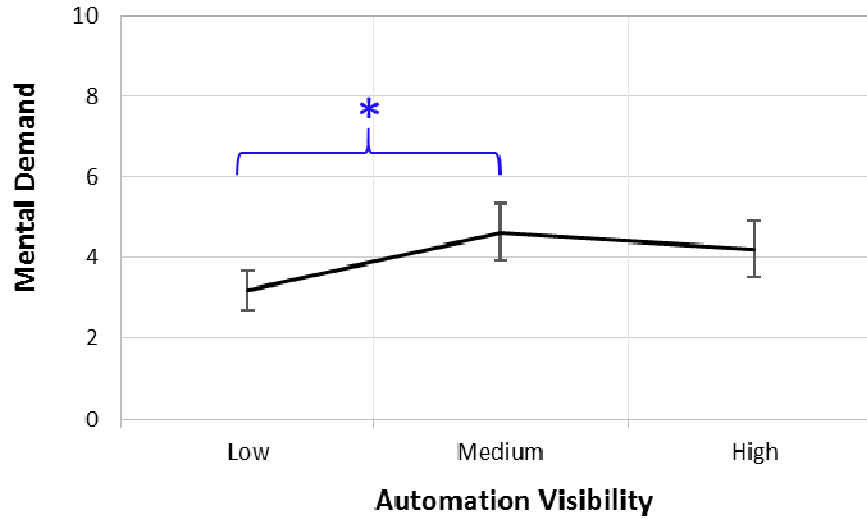
Figure 14. Mean and standard error of the total workload as assessed by the NASA-TLX.

Temporal Demand: Automation Visibility had a significant ($F_{(2,22)} = 4.56$, $p < 0.022$) impact on the temporal demand results. The paired t-test showed that the low Automation Visibility level resulted in significantly ($t_{(22)} = 2.38$, $p < 0.027$) lower temporal demand as compared to the medium level and was also marginally significantly lower ($t_{(22)} = 1.92$, $p < 0.068$) than the high Automation Visibility condition. Figure 15 shows the results of the temporal demand as a function of Automation Visibility.
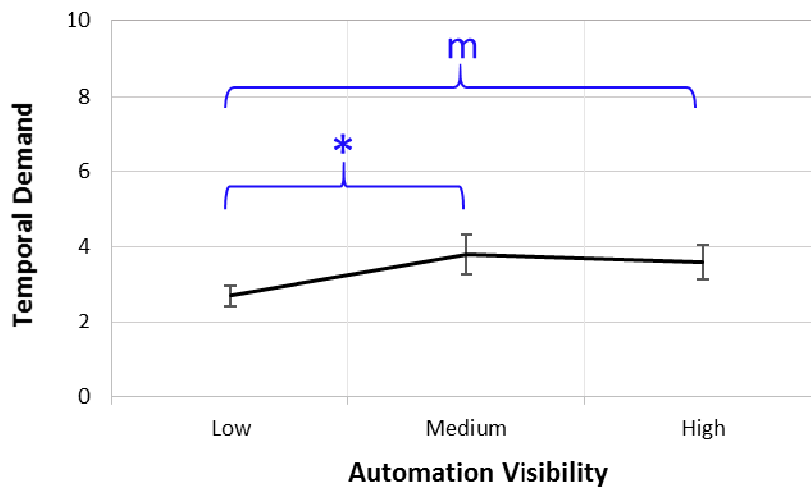

Figure 15. NASA-TLX measure of temporal demand as a function of Automation Visibility.

Performance: a significant ($F_{(2,22)} = 3.45$, $p < 0.050$) result was obtained for the interaction of Information Quality and Automation Visibility. However, a Tukey HSD test

showed no practical significance when comparing each combination of interactions. Figure 16 shows the performance results for each Information Quality level as a function of Automation Visibility.
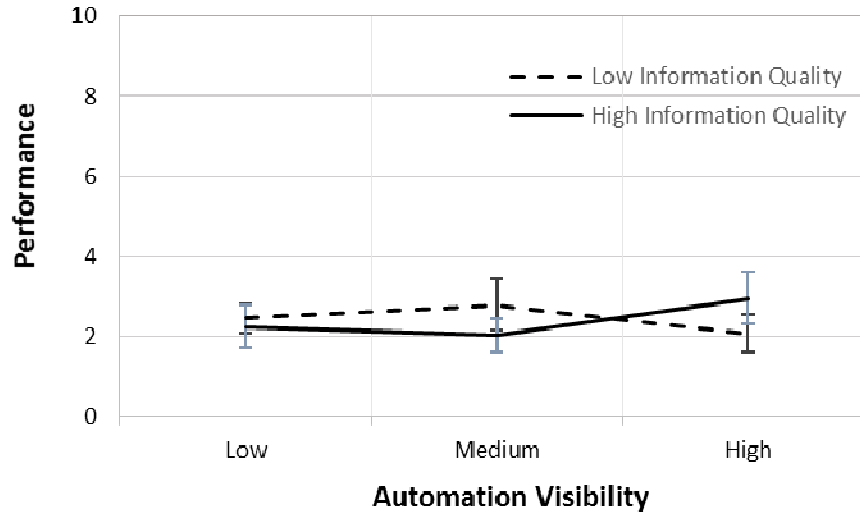

Figure 16. NASA-TLX measure of performance as a function of Automation Visibility for the low and high Information Quality levels.

6.2.2 Objective Workload

Performance on the secondary task of reporting traffic was used as an objective measure of workload. A decrease in percentage of detected targets indicates an increase in workload. Baseline measurements were also collected. Baseline measurements were taken during the first 90 seconds of the trial when participants were doing only the target detection task, before the onset of the diversion planning task.

The percentage of targets detected while participants were deciding on a diversion plan is shown in Figure 17. The baseline measurements are not shown as they were all 100%. The percentage of targets detected while participants were deciding on a diversion plan was marginally significantly greater $(F_{(1,10)} = 4.06, p < 0.072)$ in the high Information Quality condition.
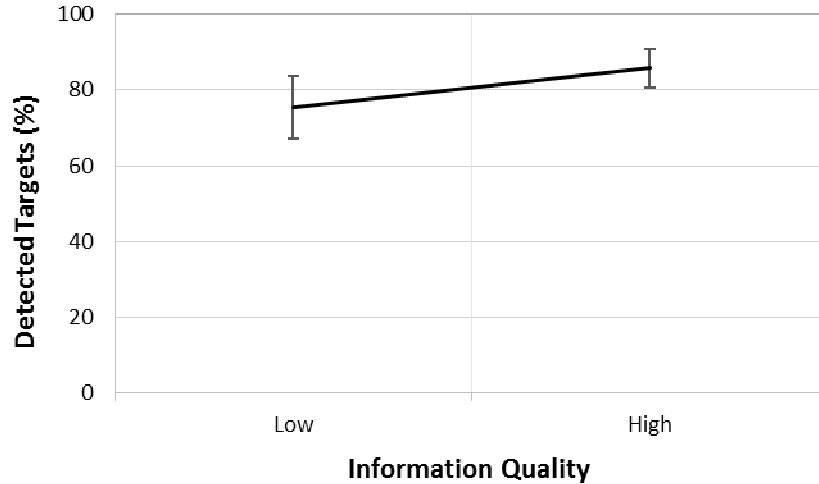
Figure 17. Percentage of targets detected during the diversion selection task.

When participants noticed that there is a discrepancy in the information displayed by the automation, it takes more effort to assess the recommendations that it provides. This result follows the trend in the time to make a diversion plan selection, shown previously in Figure 13.

6.3 Attention Allocation

Head tracking data was collected in order to capture the percentage of time participants spent looking at the aid vs. the time spent looking out the window. This measure can be used to compare the attentional requirements between conditions while the task is being completed. The differences between low Automation Visibility and the medium and high Automation Visibility levels were significant ($t_{(18)} = 4.02$, $p < 0.0008$; and $t_{(18)} = 3.24$, $p < 0.0045$, respectively). Figure 18 shows the percent time on the Diversion Aid as a function of Automation Visibility.
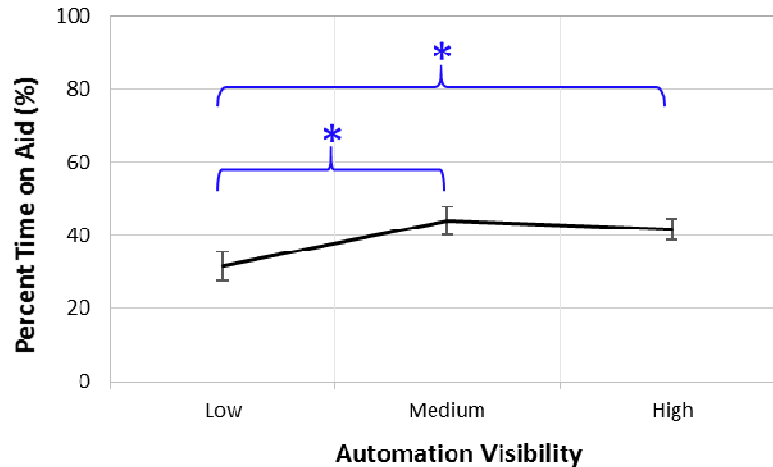
Figure 18. Percentage of time participants spent looking at the Diversion Aid while making their diversion plan selection as a function of Automation Visibility.

Since there is less information in the low Automation Visibility level, less attention is required to observe and orient to the task. Between the two higher Automation Visibility levels, the attentional requirements are similar. Although more information is provided at the high level, it is information that is relevant to the decision task and having it readily available may offload cognitive resource requirements, thus balancing the overall attentional requirements.

Information Quality also had significant ($F_{(1,9)} = 5.45$, $p < 0.044$) impact on attention allocation. Figure 19 illustrates these results, where the lower Information Quality level took more attention than the high Information Quality level.
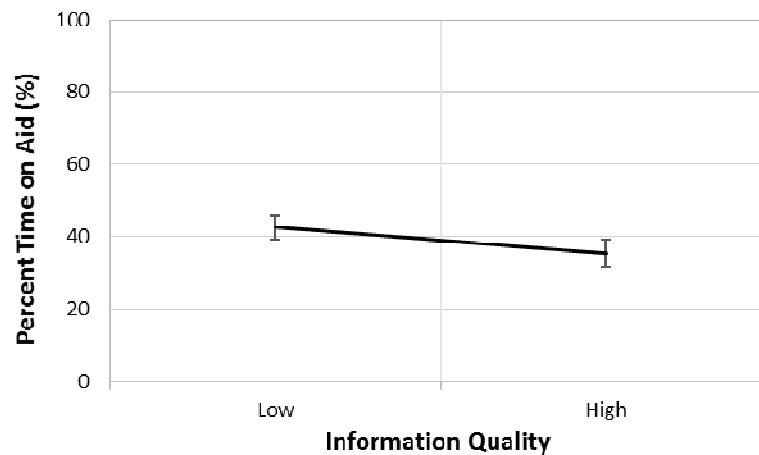

Figure 19. Percentage of time participants spent looking at the Diversion Aid while making their diversion plan selection as a function of Information Quality.

6.4   Confidence

Confidence was a self-assessment rating gathered from participants following completion of each trial. Information Quality had marginally significant ($F_{(1,11)} = 4.125, p < 0.067$) impact on confidence, with the low Information Quality condition resulting in lower confidence in the selection made. Figure 20 shows the confidence results as a function of Information Quality. As Information Quality degrades, the confidence participants have in making decisions based on that information also decreases.
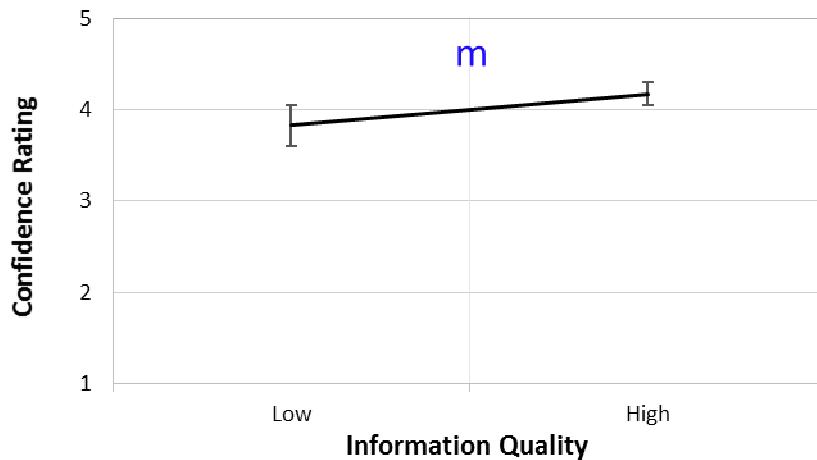
Figure 20. Confidence ratings as a function of Information Quality.

6.5   Automation Awareness

Automation awareness was a self-assessment rating gathered from participants following completion of each trial. The question asked participants to provide their level of understanding of how the Diversion Aid arrived at its recommendations. Automation Visibility was a significant ($F_{(2,22)} = 5.08, p < 0.015$) factor in this measure, Information Quality was marginally significant ($F_{(1,11)} = 4.11, p < 0.067$), and the interaction of these two independent variables was also marginally significant ($F_{(2,22)} = 2.66, p < 0.093$). Figure 21 shows the results of automation awareness as a function of Automation Visibility level. The difference between medium and high Automation Visibility was significant ($t_{(22)} = 2.69, p < 0.013$) and the difference between the low and high Automation Visibility levels was marginally significant ($t_{(22)} = 1.69, p < 0.10$).
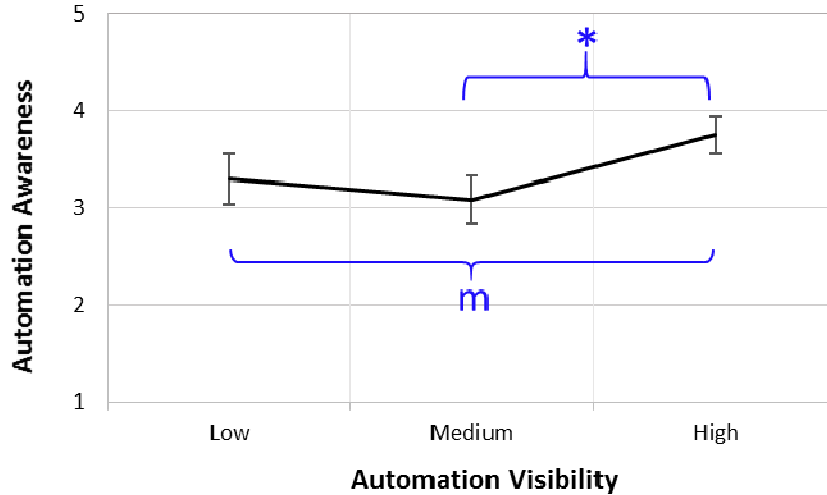
Figure 21. Automation awareness ratings as a function of Automation Visibility.

Although the low Automation Visibility level only provided one diversion plan option, participants rated their understanding of its logic closer to that of the high Automation Visibility level than the medium level. Having only one option presented meant that participants only had to understand one plan, rather than having to understand three plans. With the costs included in the high Automation Visibility level, the details of the logic are much more readily available.

There was marginal significance ($F_{(1,11)} = 4.11$, $p < 0.067$) between the low and high Information Quality results (see Figure 22).
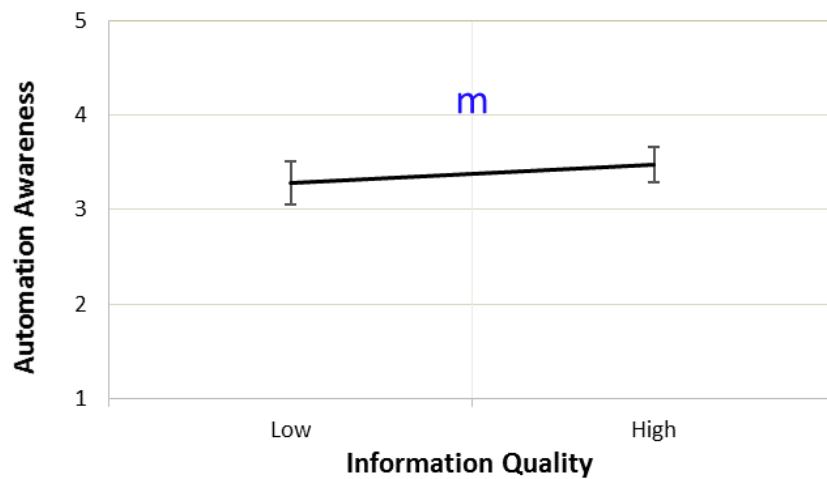
Figure 22. Automation awareness ratings as a function of Information Quality.

6.6 Trust

Trust was a self-assessment rating gathered from participants following completion of each trial. Automation Visibility, Information Quality, and their interactions all had significant ($F_{(2,20)} = 4.18$, $p < 0.030$; $F_{(1,10)} = 6.26$, $p < 0.031$; $F_{(2,20)} = 4.15$, $p < 0.031$, respectively) impact on the trust measure. For Automation Visibility, the difference in trust between the low and high levels was significant ($t_{(20)} = 2.40$, $p < 0.026$). The results of the trust ratings are shown in Figure 23 as a function of Automation Visibility.
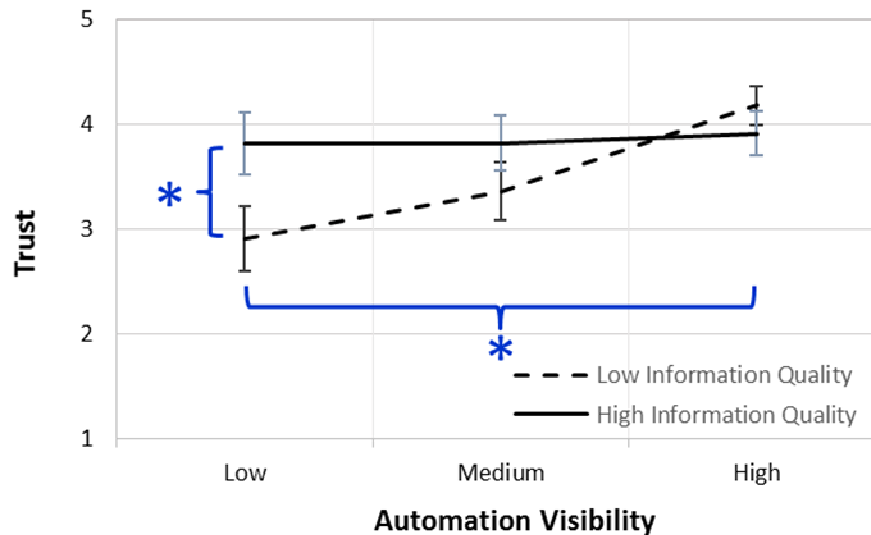
Figure 23. Trust ratings as a function of Automation Visibility.

Trust in the high Information Quality condition was the same across all three Automation Visibility conditions. In the low Information Quality condition, trust was lower than in the high Information Quality condition for both the low and medium Automation Visibility conditions. In the high Automation Visibility condition, trust in the system was the same for all three Automation Visibility conditions. Thus only when the system provides maximum information on its reasoning did the participants' level of trust in low Information Quality situations approach the (constant) level of trust in the high Information Quality situation.

6.7 Diversion Aid Display Attributes

After completing the six trials, participants filled out a post-experiment questionnaire to assess their opinions about the three Automation Visibility levels for their Display Mode (since Display Mode was a between-subjects variable). Participants were asked to distribute a total of 100 points to the three Automation Visibility levels for each of five attributes: 1) clarity of information, 2) completeness of information, 3) ease of finding information, 4) helpfulness in making a decision, and 5) preference. The higher the points assigned, the more that Automation Visibility level was preferred over the other two.

The mean scores for each attribute and Automation Visibility level are shown in Figure 24 with the standard errors shown in parentheses below each mean. On average, participants felt that low Automation Visibility was clear and easy to use, while high Automation Visibility was complex. Participants felt that the Automation Visibility levels were about equally helpful, with a slight preference for the low Automation Visibility condition. No single Automation Visibility level was clearly preferred over the others.
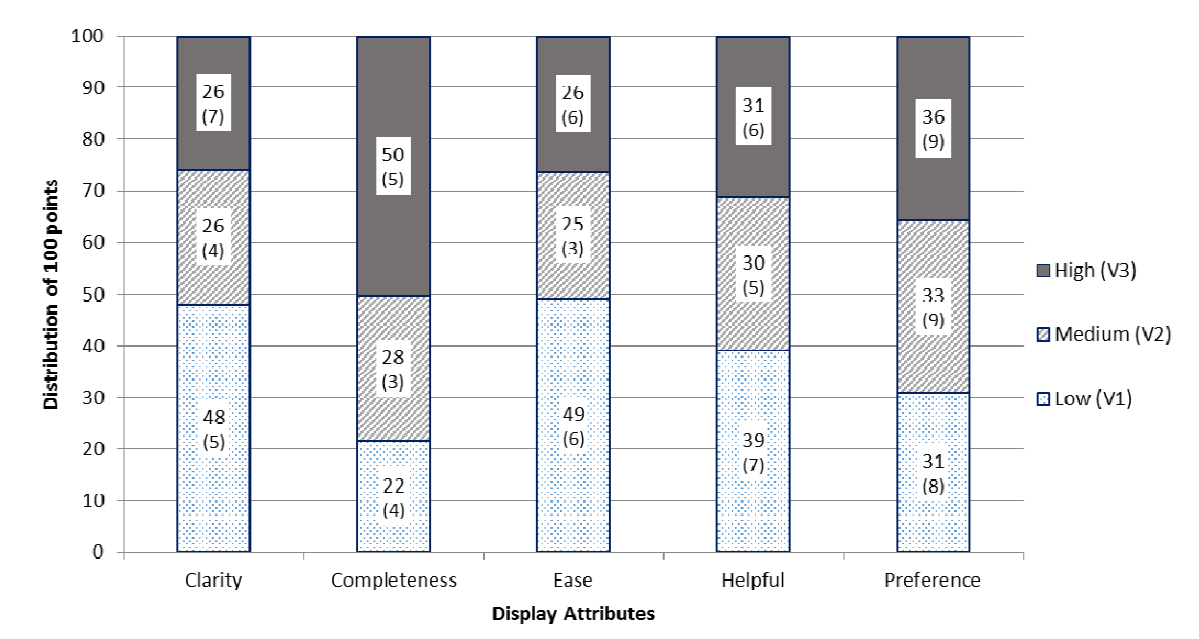
Figure 24. Mean (standard error) Diversion Aid attribute scores vs. Automation Visibility level.

Considering individual Preference ratings, it becomes clear that participants had strong preferences that varied considerably. Nine of 12 participants gave 60 points or more to

a single Automation Visibility level, but they did not all agree on which level was preferred. Figure 25 graphically shows the individual scores for the participant's preference attribute in order to convey the variety of the responses.
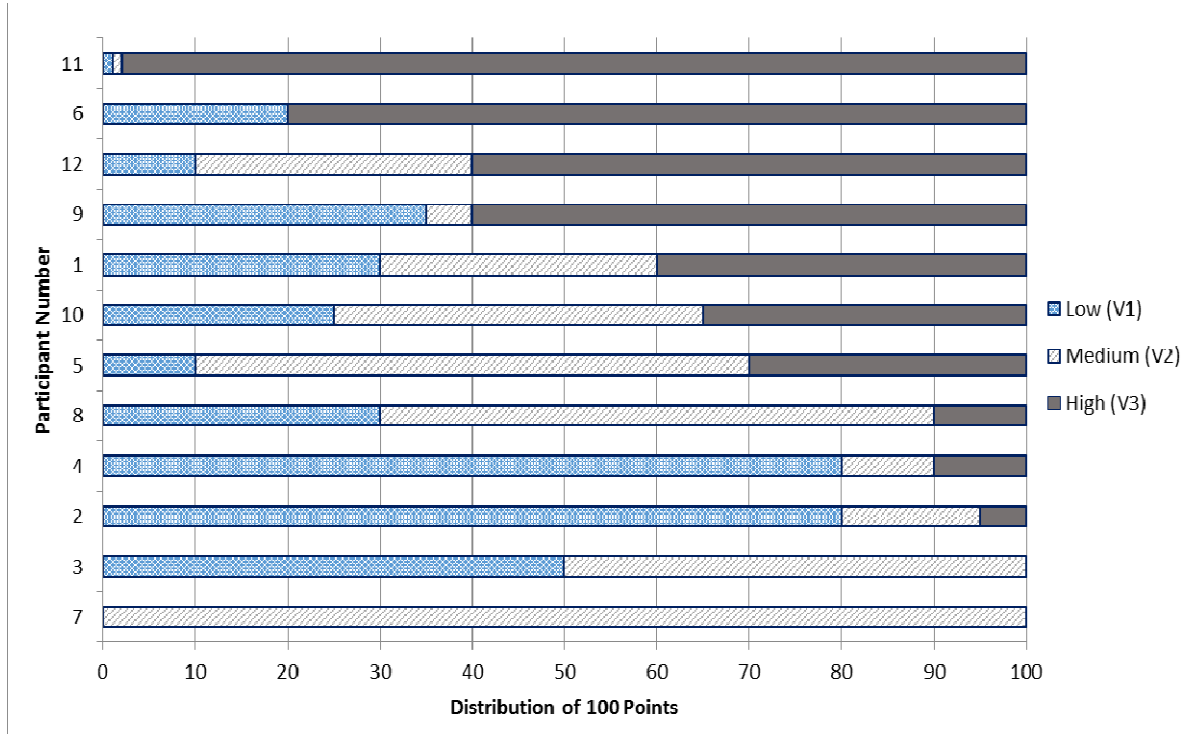
Figure 25. Individual preferences for Automation Visibility levels.

6.3.1 Decision-making Strategies

Participants were asked to describe their decision-making strategies in an open-ended question after all the trials had been completed. The strategies that participants used to make their selection were varied and depended on the Automation Visibility level. For the low Automation Visibility level, half of them trusted what the automation told them and selected the plan given as long as it seemed to make sense and they thought it was safe. For the medium Automation Visibility level, five participants adopted some level of accounting to understand why the plans were ranked as they were. Four participants did not specify a strategy, while three indicated that they looked for the option to hold and selected it because it would keep them going to the same airports. One participant simply trusted the top plan and, as long as he felt it was safe, selected it (despite being briefed that all plans were safe).

For the high Automation Visibility level, three participants indicated they were reluctant to trust the costs provided to them. One, despite having been trained on the purpose of the aid, commented that he did not care about these costs, as they were related to issues outside of his primary responsibility of getting passengers to their destinations.

6.3.2   Decision Aid Features

Participants were asked about likes and dislikes about each Automation Visibility level. The most common feedback regarding features they liked about the low and medium Automation Visibility levels were their simplicity. However, in both the low and medium Automation Visibility levels, participants said they wanted more information and reasoning behind the best plan they were being shown (i.e., higher Automation Visibility). The feedback regarding the medium Automation Visibility level was the most varied. Three participants very much liked that three options were offered to them without reasoning information (i.e., costs) to evaluate on their own, while three others commented that they thought this was the worst level to work with because they wanted to either have the best option only (i.e., low Automation Visibility) or the costs provided (i.e., high Automation Visibility). Generally, participants preferred the inclusion of reasoning information in the high Automation Visibility condition, where three pilots commented that they liked having some insight into the financial impact of their diversion decisions. Two participants, however, commented that they did not care at all about those details.

CHAPTER 7

DISCUSSION

This study identified and experimentally investigated three characteristics of information automation on the flight deck: Information Quality, Automation Visibility, and Display Complexity. Several hypotheses were tested. Each of the hypotheses is reviewed in the next section, followed by an overall discussion of the study's findings. The chapter closes with a set of recommendations generated by aggregation of the study's conclusions.

7.1   Review of Hypothesis Tests

1. Increased Automation Visibility will result in increased primary task performance, increased confidence in decisions, and increased trust in automation, but at a cost of higher workload.

    In this study, there were no performance effects due to Automation Visibility. For all conditions, the increased Automation Visibility from low to high came at the cost of higher workload and increased selection time. In low Information Quality, an increase in Automation Visibility from low to high also showed an increase in trust, eventually reaching the trust level seen at high Information Quality, where trust remained constant between Automation Visibility levels. Automation awareness was greatest in high Automation Visibility. Finally, for low Information Quality situations, confidence in automation increased between low and high Automation Visibility; however, there was a drop in confidence in the high Information Quality condition. Coupled with the automation awareness results, this suggests that confidence in their choice and automation awareness increase when pilots understand the limits of the automation, but that confidence is negatively impacted by high workload.

2. Higher Information Quality will result in better primary task performance when compared to lower Information Quality.

    Diversion plan selection performance was significantly higher when Information Quality was high when compared to selection performance when Information Quality was high when compared to selection performance when Information Quality was low.

3. Higher Automation Visibility will result in increased ability for pilots to compensate for poor Information Quality in the automation to maintain overall primary task performance (i.e., the difference in primary task performance between the low and high Information Quality conditions will be greater when Automation Visibility is low than when the visibility is high).

Making the correct diversion decision under the low Information Quality condition required participants to use information received from another source (the briefing from the confederate) to check the information from the Diversion Aid. Participants were able to compensate for poor Information Quality on average 36% of the time. Automation Visibility level did not have an effect on these results.

While previous research in this area (Sarter & Woods, 1992 and 1994b; Pritchett & Vándor, 2001; Skjerve & Skraaning, 2004; Seong & Bisantz, 2008; Bass et al., 2013) suggests that pilots should be able to compensate for poor automation decisions (in this case driven by poor Information Quality), the results of this study indicate several other factors contributed to the generally poor performance: workload, display complexity, trust, and operational biases. The complexity of the display made it difficult for participants to detect missing information, even when they knew they were looking for it (e.g., the participant who was actively searching to make sure that the unaccompanied minor was in the plan, yet failed to detect that that piece of information was missing). Their generally high trust in the automation coupled with the time pressure of the situation also caused pilots to spend less time checking for missing information. Even when they intuitively knew something was "not quite right" (as evidenced in the increased time spent making a decision in low Information Quality conditions), they often failed to detect the missing information. Another factor that affected performance overall was participants' preference for the hold option, despite the policy cost values.

4. The graphical display will result in increased task performance, increased detection of information quality issues, lower complexity, and lower workload when compared with the text display.

The research hypothesis was not supported. Selection of the best option was not better with the graphic display than with the text display, nor was there improvement in the time to

make a selection. The graphic display also did not show any improvement over the text display when looking into the differences in performance due to Information Quality issues. Complexity ratings and workload results showed no significant reduction due to the Display Mode variable. The results suggest that, for schedule information, the tabular nature of the text display supported the overall task better than the spatial display of the schedule in the graphic mode. This is an example of a situation in which competing display principles need to be assessed to determine which is more important to overall performance and workload.

7.2   Overall Discussion

The experimental investigation into the human performance impacts of three information automation characteristics of Information Quality, Automation Visibility, and Display Complexity has provided some conclusive results and others that merit further investigation. A summary of the findings are:

- Poor Information Quality was difficult for participants to detect, even when they were presented with the highest Automation Visibility level. Participants were able to compensate for poor information quality on average only about a third of the time. In the times that they did not successfully compensate, participants tended to over-trust the automation, so when information was missing and they were under high workload, they tended to choose the top plan suggested by the automation even though it was not the truly best plan according to the company policy statements.

- The level of Automation Visibility affected decision time, with low Automation Visibility leading to the fastest decision. Automation Visibility also affected workload (but not in a strictly monotonically increasing capacity). That is, the highest level of Automation Visibility did not necessarily yield the lowest workload. General consensus from the participants' qualitative responses, however, indicated that if multiple options are presented, they want some way to assess those options and understand the automation's reasoning (high Automation Visibility level).Trust in automation is affected by Information Quality, but can be compensated for by increased Automation Visibility. In low Information Quality situations, trust was lower than in high Information Quality situations for low and

medium Automation Visibility; however trust was the same at high Automation Visibility.

- A high level of trust in automation can lead to reluctance to override automation's recommendations. This has a negative impact on decision performance when Information Quality is low.

- In decision-making tasks, providing a ranked list of options without giving the reasoning behind the order results in higher workload. Providing more options in a decision-making task should only be done if the logic behind those options is also provided.

- Higher Information Quality results in lower workload.

- As Information Quality degrades, the confidence participants have in making decisions based on that information also decreases.

- As operators are exposed to more of the automation's logic, the more they trust it.

7.3   Recommendations for Design

A set of initial recommendations for design of information automation on the flight deck was provided in Honeywell's Phase I report (Rogers et al., 2013). Part of the goal of this study was to help refine and update those recommendations, and to generate new ones based on the results of the experimental investigation. The resulting recommendations presented here are organized by the three information automation characteristics studied: *Information Quality*, *Automation Visibility*, and *Display Complexity*. A complete set of the combined Honeywell and Iowa State University recommendations can be found in (Rogers et al., 2014).

7.3.1   Information Quality

1. **Appropriate levels of information quality should be defined for information automation systems, depending on the potential impact of the information on flight safety.**

   Various properties of information quality should be considered, including: intrinsic quality, contextual quality, representational quality, and accessibility. The Diversion

Aid evaluation showed that performance can be impacted by the quality of information, and depending on the specific design of the information automation system, pilots may or may not be able to compensate. Thus, it is important that the automation system meets minimum standards for information quality.

2. **Information automation systems should check for input discrepancies.**

Information automation systems that are capable of using and processing redundant sources of data or information could provide comparisons of those sources. Any discrepancies or inconsistencies identified could be annunciated to support pilot awareness.

3. **Information automation systems that produce outputs that vary in quality (e.g., accuracy, completeness, timeliness) should annunciate those variations if possible.**

Systems can be designed so that they produce partial outputs or outputs based on partial inputs (e.g., a flight path Estimated Time of Arrival that does not consider winds aloft). This might be beneficial, for example, for a decision aid or a system which performs calculations where some input parameters have minor effects on the outcomes. But the results of the study presented here indicate that pilots have difficulty in determining if there is missing information, so it may be useful to present incomplete information with supporting information about the quality (e.g., annunciation that a certain factor is not included in a calculation). Further, information automation may produce outputs that are dynamic or can become "stale," or which are inherently uncertain or probabilistic in nature, and an indication of these aspects of quality may be useful as well. For example, information that is 60 seconds old may be "real-time" in some systems (e.g. weather display) and "stale" in other systems (e.g. traffic alerting system). Some indication of the freshness or time last updated allows the pilots to bring in their understanding of the current context to decide how timely the data are.

In general, pilots in this study found it difficult to compensate for poor information quality, most likely due to factors such as interface design, task difficulty, and display complexity. The more the information automation system can display its own assessment of information quality, the more redundancy it provides in the joint human-automation system, since both the pilot and the automation should ideally be assessing the quality of the information.

4. **Training on information automation should consider rules of thumb for how to assess the quality of information outputs.**

As information automation systems become more powerful (e.g., adaptive systems that can assess contextual factors, intelligent systems that reason and learn), it may be more important for pilots to receive specific training on how these systems work, what their limitations are, how to verify their outputs, and so on. Further, as information automation supports pilots more and more in management and decision making tasks, it would be useful to train the best ways to utilize the aids to support those tasks.

5. **For effective usage of information automation systems and their output, training should be provided on issues such as information quality, distractions, workload, over-trust, and skill degradation.**

Information quality as defined in this work goes beyond accuracy and precision. In cases where the pilots are responsible for monitoring the outputs of an information automation system, they need strategies for searching for and detecting information quality issues. For instance, when assessing the diversion plans in our empirical evaluation, pilots may have benefited from being trained on a strategy to check through categories of information to ensure that they could identify missing information.

7.3.2   Automation Visibility

1. **Information for verifying or checking system reasoning and output should be available, easy to detect, and easy to access. It should be made obvious if some information that is normally presented is missing.**

   The complexity of the task, the design of the interface, and the saliency of the information all play a role in whether pilots can detect that something is missing or inaccurate. Even in cases of high automation visibility, where the automation reveals its reasoning to the pilot, it is often difficult to notice what is not there. Thus the interface should provide support to help pilots know what information to look for to assist in cases where that information is missing. Explanations of system behavior and states, and quality of information outputs should be available upon demand. The results this study indicate such information led to increased automation awareness and to information automation systems that were more preferred by the pilots.

2. **Presentation of information to help pilots understand information automation state and outputs should be balanced against potential increases in pilot workload due to the time and attention needed to process this extra information.**

   Even though information automation outputs are usually beneficial, if they require an inordinate amount of workload to validate (e.g., manual searching and integrating of information), the costs could outweigh the benefits from a human performance perspective. A balance between having automation visibility information and the time and effort needed for the pilot to process that information is important. In some cases, a small amount of automation visibility information, or automation visibility information that can be accessed on demand but not presented automatically, should be considered. If visibility information can be built into the information automation outputs themselves, less processing may be required to validate the outputs. The results the empirical study suggest that pilots may not spend extra effort searching for validation information.

3. **If an Information Automation system provides choices or alternatives, information on how those choices were determined and their relative merits should be provided.**

    The evaluation did not show a clear preference for a decision aid that showed the best option only versus one that showed multiple options with cost information. However, most participants wanted visibility into how those options were determined. They felt it was too much work to try and figure out why the system prioritized the options the way that it did. In comparison to the best option or options with supporting information, presenting options with no supporting information resulted in lower performance, slower performance, higher workload, more attention, and lower automation awareness.

7.3.3 Display Complexity

1. **Information automation display complexity can compromise usability – in some cases it may be better to have a less capable system that reduces complexity and is easier to use.**

    Adding new functions to an existing display are often seen as a way to improve operational safety and efficiency. Each additional function can add to the complexity of a single system or device in terms of pilots' understanding of its behavior and the ease of interacting with the device. This could negatively affect user workload and the overall usability of the system.

CHAPTER 8

CONCLUSION

8.1   Summary of Research

Based on a previously published definition and framework of information automation on the flight deck (Rogers et al., 2013), both heuristic and analytical methods were employed to generate and refine a set of characteristics to describe information automation in this domain. Three of these characteristics were selected for further experimental study into the human performance impacts of flight deck information automation. Analysis of the experimental results informed design recommendations to address the observed impacts.

8.2   Contributions

This work addresses a previously identified need for a more formal definition and characterization of flight deck information automation (FAA, 2013b). As more information becomes available on the flight deck, it is crucial that the human performance impacts of information automation systems be well understood by designers so that pilots and the automation are able to work in harmony to ensure mission safety as well as a more efficient flying environment as envisioned by NextGen. Lessons learned by past accidents have shown that discord between pilots and automation can have catastrophic consequences (e.g., Asiana Airlines, 2014; et d'Analyses, 2009; Palmer, 2013). With the increasing amount of information available to pilots, information automation is seen to be equally as important as control automation to achieve these safety and efficiency goals.

This work also provides the experimental results and analyses that informed a first set of recommendations for the design of flight deck information automation systems. With these recommendations, the human factors issues associated with these systems can be addressed.

Lastly, by stepping through each stage of the process for developing an experiment to test a subset of the characteristics identified, the work provides a roadmap for developing further experimental studies to expand on the results and recommendations provided here. This expansion of research will be necessary in order to further the understanding of the human factors impacts of information automation on the flight deck.

8.3 Future Work

This work was a first step in understanding the human factors impacts of flight deck information automation systems, but there are open questions that warrant further investigation. For example, this study only looked at a subset of the dimensions of information quality (data that was missing or incomplete). Furthermore, the human factors impacts of the characteristics that were not studied experimentally during this research are also important to address. Additional recommendations for design could then be generated to complement those created from this work.

Another impact of information automation that would be important to understand for mitigation purposes is the area of cognitive skill degradation. Designers are continuing to improve the capabilities of the information automation technology available to pilots, but if and when something goes wrong, will pilots be able to take over those tasks that have been done for them? How often should pilots receive training to ensure they are not losing important skills to accomplish the tasks that have been taken over by automation? These are just a few of the questions regarding cognitive skill degradation that will need to be addressed as information automation becomes more sophisticated and capable.

REFERENCES

Asiana Airlines Inc. (2014). Accident Investigation Submission. *Available*: http://www.scribd.com/doc/215600314/Asiana-Airlines-Accident-Investigation-Submission (*accessed, May 31, 2014*).

Adelman, L., Christian, M., Gualtieri, J., & Johnson, K. L. (1998). Examining the effects of cognitive consistency between training and displays. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 28(1), 1-16.

Andre, A. D., & Cutler, H. A. (1998). Displaying uncertainty in advanced navigation systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(1), 31-35. SAGE Publications.

Bass, E. J., Baumgart, L. A., & Shepley, K. K. (2013). The effect of information analysis automation display content on human judgment performance in noisy environments. *Journal of cognitive engineering and decision making*, 7(1), 49-65.

Bass, E. J., & Pritchett, A. R. (2008). Human-automated judge learning: A methodology for examining human interaction with information analysis automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(4), 759-776.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, 41(3), 16.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: defining customer-centered systems*. Elsevier.

Billings, C. E. (1991). Toward a human-centered aircraft automation philosophy. *The International journal of aviation psychology*, 1(4), 261-270.

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

Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. *Human performance in automated systems: Current research and trends*, 264-269.

Bisantz, A. M., Marsiglio, S. S., & Munch, J. (2005). Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(4), 777-796.

Boy, G. A. (2008). *Human-centered development of perceived complexity criteria: Developed criteria*. Technical Report DGAC/EURISCO.

Boyd, J. (1987). A discourse on winning and losing. Maxwell Air Force Base, AL: Air University Library Document No. M-U 43947 (Briefing slides).

Cummings, M. L., Sasangohar, F., Thornburg, K. M., Xing, J., & D'Agostino, A. (2010). Human-system interface complexity and opacity part i: literature review. *Massachusetts: Institute of Technology Editors*.

de Winter, J. C. F., & Dodou, D. (2011). Why the Fitts list has persisted throughout the history of function allocation. Cognition, Technology & Work, 1-11.

Degani, A., Barshi, I., & Shafto, M. G. (2013). Information Organization in the Airline Cockpit Lessons From Flight 236. *Journal of Cognitive Engineering and Decision Making*, 1555343413492983.

Deveans, T., & Kewley, R. (2009). *Overcoming information overload in the cockpit* (ORCEN Tech. Rep. No. DSE-TR-0904). West Point, NY: U.S. Army.

Dix, A., Finlay, J. E., Abowd, G. D., & Beale, R. (Eds.). (2004). *Human-Computer Interaction*. Pearson Education.

Dorneich, M. C., McGrath, K. A., Dudley, R. F., & Morris, M. D. (2013, October). Analysis of the Characteristics of Adaptive Systems. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (p. 888-893). IEEE.

Dorneich, M. C., Whitlow, S. D., Miller, C. A., & Allen, J. A. (2004). A superior tool for airline operations. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 12(2), 18-23.

Endsley, M. R. (1996). Automation and situation awareness. *Automation and human performance: Theory and applications*, 163-181.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Endsley, M.R. (2000). Theoretical Underpinning of Situation Awareness: Critical Review in Mica R. Endsley and Daniel J. Garland (Eds.) *Situation awareness analysis and measurement*. Lawrence Erlbaum Associates, Mahwah, New Jersey. 3-32.

Endsley, M. R. (2010). Situation awareness in aviation systems. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Handbook of aviation human factors* (pp. 12-1-12-22). Boca Raton, FL: CRC Press.

English, L. P. (2009). *Information quality applied: best practices for improving business information, processes and systems*. Wiley Publishing.

et d'Analyses, B. D. E. (2012). Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro– Paris.

FAA. (2013a). NextGen Implementation Plan. *Available*: http://www.faa.gov/nextgen/implementation/media/NextGen_Implementation_Plan_2 013.pdf (*accessed August 13, 2013*).

FAA (2013b). "Operational Use of Flight Path Management Systems", Final Report of the Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team, Flight Deck Automation Working Group, September 5, 2013. *Available*: http://www.faa.gov/about/office_org/headquarters_offices/avs/offices/afs/afs400/parc /parc_reco/media/2013/130908_PARC_FltDAWG_Final_Report_Recommendations. pdf (*accessed May 9, 2014*).

Fadden, D.M. (1990). Aircraft automation challenges. In *Abstracts of AIAA-NASA-FAA-HFS Symposium, Challenges in Aviation Human Factors: The National Plan*.

Feigh, K.M., Dorneich, M.C., & Hayes, C.C. (2012). Toward a Characterization of Adaptive Systems A Framework for Researchers and System Designers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(6), 1008-1024.

Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system*. National Research Council, Washington, DC.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & and Owen, G. Flight Deck Automation Issues, *International Journal of Aviation Psychology*, 9(2), 1999, pp. 109-123.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Helmreich, R. L., & Musson, D. M. (2000). Threat and error management model: Components and examples. *British Medical Journal*.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(4), 730-741.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an" aid" can (and should) go unused. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(2), 221-242.

Landry, S. J. (2009). Flight Deck Automation. In *Springer Handbook of Automation* (pp. 1215-1239). Springer Berlin Heidelberg.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80.

Myers, B. L., Kappelman, L. A., & Prybutok, V. R. (1997). A comprehensive model for assessing the quality and productivity of the information systems function: toward a theory for information systems assessment. *Information Resources Management Journal (IRMJ)*, 10(1), 6-26.

Mosier, K. L., Fischer, U., Morrow, D., Feigh, K. M., Durso, F. T., Sullivan, K., & Pop, V. (2013). Automation, Task, and Context Features Impacts on Pilots' Judgments of Human–Automation Interaction. *Journal of Cognitive Engineering and Decision Making*, 1555343413487178.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Basic Books.

Palmer, B. (2013). *Understanding Air France 447*. William Palmer.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6), 695-706.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(3), 286-297.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.

Pritchett, A. R., & Vándor, B. (2001). Designing situation displays to promote conformance to automatic alerts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(4), 311-315. SAGE Publications.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics, IEEE Transactions on*, (3), 257-266.

Reeves, C. A., & Bednar, D. A. (1994). Defining quality: alternatives and implications. *Academy of management Review*, 19(3), 419-445.

Rogers, W., Whitlow, S., Letsu-Dake, E., Ott, J., & Dorneich, M. (2013). Flight Deck Information Automation: Analysis and Recommendations: Phase I report. *Honeywell Phase I report,* Contract OTA DTFAWA-10-A-80031, Task 10 to the Human Factors Research and Engineering Group Federal Aviation Administration.

Rogers, W., Letsu-Dake, E., Dorneich, M., Dudley, R., Nelson, E., Dillard, M., Whitlow, S., Almquist, J., Hamblin, C., & Ott, J. (2014). Flight Deck Information Automation: Analysis and Recommendations. *Honeywell Final Report,* Contract OTA DTFAWA-10-A-80031, Task 10 to the Human Factors Research and Engineering Group Federal Aviation Administration.

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. (1994a). Decomposing automation: Autonomy, authority, observability and perceived animacy. In *First Automation Technology and Human Performance Conference* (p. 22-26).

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

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

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(4), 553-569.

Sarter, N. B., & Woods, D. D. (2000). Team play with a powerful and independent agent: a full-mission simulation study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(3), 390-402.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, 2, 1926-1943.

Seagull, F. J., & Sanderson, P. M. (2001). Anesthesia alarms in context: An observational study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1), 66-78.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37–63). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Seong, Y., & Bisantz, A. M. (2002). Judgment and trust in conjunction with automated decision aids: A theoretical model and empirical investigation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(3), 423-427. SAGE Publications.

Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, 38(7), 608-625.

Sheridan, T. B. (1980). Computer control and human alienation. *Technology review*, 83(1), 65-73.

Sheridan, T. B. (2001). Rumination on automation, 1998. *Annual Reviews in Control*, 25, 89-97.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB.

Skjerve, A. B. M., & Skraaning Jr, G. (2004). The quality of human-automation cooperation in human-system interface for nuclear power plants. *International journal of human-computer studies*, 61(5), 649-677.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, 58(12), 1720-1733.

Tenney, Y. J., Rogers, W. H., & Pew, R.W. (1998). Pilot Opinions on Cockpit Automation Issues. *International Journal of Aviation Psychology*, 8(2), pp. 103-120.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5-33.

Whitlow, S. D., Dorneich, M. C., Funk, H. B., & Miller, C. A. (2002). Providing appropriate situation awareness within a mixed-initiative control system. In *Systems, Man and Cybernetics, 2002 IEEE International Conference on* (Vol. 5). IEEE.

Wickens, C. D. (1994). Designing for situation awareness and trust in automation. *Proceedings of IFAC integrated systems engineering*, 77-82.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current directions in psychological science*, *11*(4), 128-133.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (Eds.) (1997). *Flight to the future: Human factors in air traffic control*. Washington, DC: National Academy Press.

Wickens, C.D., Mavor, A., Parasuraman, R., & McGee, J.M. (1998). *The future of air traffic control: Human operators and automation*. Washington, DC: National Academy Press.

Wiener, E. (1989). Human factors of advanced technology (glass cockpit) transport aircraft. (NASA Contractor Report No. 177528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. *Automation and human performance: Theory and applications*, 3-17.

Xing, J. (2007). *Information complexity in air traffic control displays* (p. 797-806). Springer Berlin Heidelberg.

Xing, J. (2008). *Designing questionnaires for controlling and managing information complexity in visual displays* (No. DOT/FAA/AM-08/18). Federal Aviation Administration, Oklahoma City, OK, Civil Aerospace Medical Inst.

APPENDIX A

POST-TRIAL QUESTIONNAIRE

1. How confident were you in your decision?

1	2	3	4	5
Not at all	Slightly	Somewhat	Fairly	Very

2. To what degree do you understand how the aid came up with its recommendations?

1	2	3	4	5
No understanding	Slight understanding	Some understanding	Fairly Good understanding	Very High understanding

3. How much did you trust the recommendation(s) given to you?

1	2	3	4	5
Not at all	Slightly	Somewhat	Mostly	Very Much

4. How easy was it for you to find information on the display?
   a) I could see the information effortlessly.
   b) I could find the information with a few quick glances.
   c) I could find the information by searching in a local area of the display.
   d) I had to search through the display to find the information.

5. How easy was it for you to understand / comprehend the displayed information?
   a) The information was very straightforward. I could understand the meaning without thinking.
   b) I could integrate the pieces of information and use them properly, but would prefer that information be presented in a less intermingled manner.
   c) I needed to use some strategies to manage the displayed information. That took my mental resources away from other tasks.
   d) I had to simultaneously associate (or to relate) multiple pieces of displayed information to use the display. It was difficult to hold them all at once.

6. How would you rate the perceptual complexity of the display?
   a) The display looked simple and clear; I could find the needed information easily and quickly.
   b) The display looked busy but I could find the information with a little effort.
   c) Many pieces of information did not relate to my task; they adversely affected my perception of information.
   d) The display looked too busy for me to find the information.

7. Please explain why you made the choice you did.

APPEDNIX B

POST-EXPERIMENT QUESTIONNAIRE

1) What strategies did you use to make your decisions within each display?

    A – best plan only

    B – rank-ordered top three plans

    C – rank-ordered top three plans with decision cost

2) For each display, please list three things you liked.

    A – best plan only

    B – rank-ordered top three plans

    C – rank-ordered top three plans with decision cost

3) For each display, please list three things you would improve.

    A – best plan only

    B – rank-ordered top three plans

    C – rank-ordered top three plans with decision cost

Please distribute 100 points between the three display types for each of the following attributes (more points indicate higher preference; 100 point sum per attribute):

Attribute / Display Type	Clarity of information	Completeness of information	How easy it was to find information	Helpfulness in performing the task	Preference
A: best plan only					
B: rank-ordered top three plans					
C: rank-ordered top three plans with decision cost					
SUM must equal 100 in each column	SUM = 100	SUM = 100	SUM = 100	SUM = 100	SUM = 100

If you have any comments you would like to add regarding the table above, please write them here: