Report No. NAWCADWAR-94140-60

TASK-DEPENDENT EFFECTS OF AUTOMATION: THE ROLE OF INTERNAL MODELS IN PERFORMANCE, WORKLOAD, AND SITUATIONAL AWARENESS IN A SEMI-AUTOMATED COCKPIT

Lt. Meghan Carmody, U.S.N.
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION
P.O. Box 5152
Warminster, PA 18974-0591

1 MARCH 1994

FINAL REPORT
Period Covering December 1991 to December 1993

Approved for Public Release; Distribution is Unlimited.

Prepared for
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

NOTICES

REPORT NUMBERING SYSTEM — The numbering of technical project reports issued by the Naval Air Warfare Center, Aircraft Division, Warminster is arranged for specific identification purposes. Each number consists of the Center acronym, the calendar year in which the number was assigned, the sequence number of the report within the specific calendar year, and the official 2-digit correspondence code of the Functional Department responsible for the report. For example: Report No. NAWCADWAR-92001-60 indicates the first Center report for the year 1992 and prepared by the Air Vehicle and Crew Systems Technology Department. The numerical codes are as follows:

CODE	OFFICE OR DEPARTMENT
00	Commanding Officer, NAWCADWAR
01	Technical Director, NAWCADWAR
05	Computer Department
10	AntiSubmarine Warfare Systems Department
20	Tactical Air Systems Department
30	Warfare Systems Analysis Department
50	Mission Avionics Technology Department
60	Air Vehicle & Crew Systems Technology Department
70	Systems & Software Technology Department
80	Engineering Support Group
90	Test & Evaluation Group

PRODUCT ENDORSEMENT — The discussion or instructions concerning commercial products herein do not constitute an endorsement by the Government nor do they convey or imply the license or right to use such products.

Reviewed By: ______________________ Date: 12/22/94
Branch Head

Reviewed By: ______________________ Date: 12/22/94
Division Head

Reviewed By: T M Mitchell Date: 12/22/94
Director/Deputy Director

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
1 MARCH 1994

3. REPORT TYPE AND DATES COVERED
FINAL 12/91 – 1293

4. TITLE AND SUBTITLE
TASK-DEPENDENT EFFECTS OF AUTOMATION: THE ROLE OF INTERNAL MODELS IN PERFORMANCE, WORKLOAD, AND SITUATIONAL AWARENESS IN A SEMI-AUTOMATED COCKPIT

5. FUNDING NUMBERS

6. AUTHOR(S)
LT. MEGHAN A. CARMODY, U.S.N.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER; AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCADWAR-94140-60

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER; AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The present study investigated the effects of automating different aviation tasks on a pilot's ability to regain manual control following automation failure. The investigation employed a version of the Multi-Attribute Task (MAT) battery (Comstock and Arnegard, 1990) which presents subjects three aviation-relevant tasks: A Tracking task, a System Monitoring task, and a Fuel Management task. Specifically, this study examined task-specific effects on performance, workload, and situational awareness of removing the human operator from the control loop for long periods of time and then requiring him/her to suddenly reenter that loop. A hypothesized task distinction was formulated on the basis of the dynamic versus stable qualities of the internal cognitive model guiding the decision-making process within a particular task. This distinction is presented within the context of a theoretical model of human decision-making in complex, semi-automated cockpits. Results indicated task-specific effects of automation on human performance. Furthermore, data regarding task-specific effects of automation on situational awareness were strongly indicative of the hypothesized distinctions.

14. SUBJECT TERMS
PILOT PERFORMANCE, COCKPIT AUTOMATION, PILOT WORKLOAD, SITUATIONAL AWARENESS, INTERNAL COGNITIVE MODELS

15. NUMBER OF PAGES
456

16. PRICE CODE

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

NAWCADWAR-94140-60

TASK-DEPENDENT EFFECTS OF AUTOMATION: THE ROLE OF INTERNAL MODELS IN PERFORMANCE, WORKLOAD, AND SITUATIONAL AWARENESS IN A SEMI-AUTOMATED COCKPIT

by

MEGHAN A. CARMODY, B.S.

A DISSERTATION

IN

PSYCHOLOGY

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Chairperson of the Committee

Accepted

Dean of the Graduate School

December, 1993

ABSTRACT

The present study investigated the effects of automating different aviation tasks on a pilot's ability to regain manual control following automation failure. The investigation employed a version of the Multi-Attribute Task (MAT) Battery (Comstock and Arnegard, 1990) which presents subjects three aviation-relevant tasks: a Tracking task, a System Monitoring task, and a Fuel Management task. Specifically, this study examined task-specific effects on performance, workload, and situational awareness of removing the human operator from the control loop for long periods of time and then requiring him/her to suddenly reenter that loop. A hypothesized task distinction was formulated on the basis of the dynamic versus stable qualities of the internal cognitive model guiding the decision-making process within a particular task. This distinction is presented within the context of a theoretical model of human decision-making in complex, semi-automated cockpits.

Results indicated task-specific effects of automation on human performance. Furthermore, data regarding task-specific effects of automation on situational awareness were strongly indicative of the hypothesized distinctions.

ACKNOWLEDGEMENTS

I would like to acknowledge, with sincere appreciation, the interests, efforts, and support of the individuals and institutions that aided my completion of this document. I must first recognize the involvement of my doctoral committee, including Dr. Jeffrey Elias, Dr. Jonathan Gluckman, Dr. Mica Endsley, Dr. Robert Bell, and Dr. Mohamed Ayoub. Because I completed my doctoral work away from Texas Tech University, their roles were particularly difficult. Despite this fact, my committee members offered their fullest efforts and provided me with a diverse base of technical and professional expertise. From proposal to defense, my committee members were supportive and enthusiastic, yet they still provided the intense scrutiny necessary for the doctoral candidate to reap all benefits from the dissertation experience.

I would like to express a special appreciation for the chair of my committee, Dr. Elias. I benefitted greatly from his expansive technical skill in the area of statistics, and his willingness to devote intense effort to the dissection of my rather lengthy experimental results. I would also like to express my appreciation to Dr. Endsley, whose expertise in the area of situational awareness was invaluable to the present work.

I must also acknowledge the help and immense support of my fellow Naval Aerospace Engineering Psychologists, as well as my colleagues at the Naval Air Warfare Center, Aircraft Division, Warminster. In particular, I am indebted to the Adaptive Function Allocation (AFAIC) program, and the program head, Dr. Jeffrey Morrison. Furthermore, I would especially like to acknowledge the unique contributions of my branch head, LCDR John Schmidt and my committee member by special appointment, Dr. Jonathan Gluckman.

I am intensely grateful to LCDR John Schmidt for his patience and his moral, as well as professional, support in my completion of this dissertation. I owe a great deal to his experience and guidance.

I must reserve a distinguished acknowledgement for the help of Dr. Jonathan Gluckman, who served as an irreplaceable mentor. He pushed me and tested my limits, but always provided support. I admire his skill and creativity as an experimenter, and I will be forever grateful for his willingness to share his knowledge.

Finally, I must express my undying love and appreciation for my family, especially my mother and father. Their unyielding faith, encouragement, devotion, and love cannot possibly be captured in words, but any acknowledgements of my educational or other achievements would be ridiculously lacking without their inclusion. No one could imagine or wish for a more supportive family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

During the Second World War, scientists and technologists began to see the need for a systems approach to the interaction of people with their machines, the latter of which were becoming increasingly complex. In recent years, not only has machine complexity increased dramatically, but very often the time afforded operators to interact with their equipment has decreased. One modern task environment that exemplifies this phenomena is military aviation.

The modern aircraft pilot is placed within an environment that is dynamically challenging to his/her physical, perceptual, and cognitive abilities. The information load on pilots is very high, and the high operating speeds of most aircraft serve to increase the demands of flight-related tasks by dramatically reducing decision making time. This is particularly true in the case of emergency situations, when quick response is of the essence and cognitive load increases dramatically (Hurst, 1976).

According to Trejo, Lewis, and Blankenship (1987), within the consuming monitoring and control functions of modern aviation, "the capacity of the human to perceive, integrate, remember, and use information may be challenged" (p. v). Logical implications have lead to the conclusion that as the technological complexities of modern systems continue to expand, the corresponding increase in the informational load upon the human operator increases the probability of human error. Decreasing this load has therefore been the fundamental goal of automation implementation.

Automation

Automation is a concept with which most people in modern western civilization are familiar, yet, at the same time, it is somewhat difficult to explicitly define. This is due, in part, to varieties of opinion with respect to what automation entails, and in part to the existence of different degrees and categories of automation.

Parasuraman, Bahri, Deaton, Morrison, and Barnes (1990) refer to the definitions of several authors in their attempt to delineate the meaning of automation, including the American Heritage Dictionary's (1986) "automatic operation or control of a process, equipment, or a system" (p. 5), and Wiener and Curry's (1980) somewhat sardonic, "a collection of tyrannical self-serving machines" (p. 996). As automation can involve anything from simple mechanical devices to elaborate designs in artificial intelligence, Parasuraman et al. (1990), with respect to aviation, decide upon the operative description, "a device that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by the pilot" (p. 6).

Purpose of Automation

Weiner and Curry (1980) describe three major factors in the impetus for automation. The first concerns technology. According to the authors, "rapid improvement in performance, and decrease in size, cost, and power consumption of various electronic devices, sensors, and display media, make automation of many flight-deck (as well as ground-based) systems a reasonable alternative to traditional manual operation" (p. 997). Safety is the second area of consideration. Greater than 50% of aircraft accidents can be attributed to human error (Weiner and Curry, 1980). Additionally, automation is often implemented for economic reasons, as savings may be acquired through decreased fuel consumption and labor costs. These three major reasons for implementing automation are further broken down into eight components by Wiener (1988):

1. Availability of technology;
2. Safety;
3. Economy, reliability, and maintenance;
4. Workload reduction and certification of two-pilot transport air-craft ;
5. More precise flight maneuvers and navigation;
6. Display flexibility;
7. Economy of cockpit space; and
8. Special requirements of military missions.

Wickens (1984), too, describes three reasons for automation. These include the allocation to automation of those functions that are potentially dangerous to humans and/or those which humans cannot do; those activities humans often perform poorly due to overloading or underloading of processing capacity; and, finally, those tasks needed "to supplement or augment human perception, memory, attention, or motor skill" (p. 334).

Not only are there different purposes for automation, there are various types as well. Wickens and Kramer (1985) discuss three major types of automation that may be implemented in human-computer systems: "automation that assists", "automation that replaces", and "adaptive automation" (p. 335). The first two types (automation that assists and automation that replaces) are more traditional forms of automation in which specific functions are allocated to human and automated components in a static manner. Allocation strategy is implemented early in the design process. Adaptive automation, on the other hand, is implemented in a dynamic manner, so that the functions allocated to the human and automated components change with the changing demands and characteristics of the system (Morrison, Gluckman, and Deaton, 1991).

Although there are undoubtedly several advantages to the use of static automation, one must temper this with a consideration of the several potential disadvantages. For instance, the increasing movement of pilots into the role of what Sheridan (1987) has termed the position of "supervisory monitor," or, "control exercised through mediation of an intelligent controller....rather than by the supervisor's direct action" (Neville and Moray, 1986), has many investigators and aviators concerned.

Many of the concerns regarding the drawbacks of "over-automation," as well as its potential consequences, have been addressed by several authors (Chambers and Nagel, 1985; Parasuraman, 1987; Parasuraman, Bahri, and Molloy, 1991; Wiener, 1977; Wiener and Curry, 1977; 1980; Wiener, 1988), who seriously question the overly optimistic view that automation can virtually eradicate human error. In fact, Wiener and Curry (1977; 1980) have summarized a subset of accident reports which indicate that, under certain circumstances, such automation can increase the incidence of human error. Within the reviewed cases, the same general problem areas resurface. These deal most fundamentally with the aspects of skill degradation and inadequate monitoring, the former of which can be primarily attributed to lack of practice, and the latter of which can be primarily attributed to loss of situational awareness and human over-reliance on the automated component of the system. With respect to these subject matters, there are undoubtedly a variety of related issues which compound to formulate an interesting, albeit complex, problem area within cockpit automation designs.

Although in the past, the design philosophy concerning automation has been primarily technologically based (what can be automated?), more recent concerns suggest a concentration upon a human-centered philosophy (what should be automated, in order to enhance human performance?) (Morrison, Gluckman, and Deaton, 1991; Weiner and Curry, 1980).

Weiner and Curry (1980), for example, introduced the dimensions of Monitoring and Control, suggesting that removing too much manual control leads to vigilance decrements, while removing too little leads to problems concerning operator information overload. Somewhere along a continuum between the two dimensions, there is an optimal level of human operator control.

The more difficult question, therefore, requires a deeper understanding of human information processing behaviors and abilities within an automated environment. Continued research efforts must be directed toward a resolution of this problem if automation is to successfully decrease the information load upon the operator and increase overall system performance. After all, one cannot successfully develop automated aids to human information processing without first attempting to understand information processing in the presence of such aids. It is the interaction of the elements involved in information processing (both human and computer) that creates the intrigue surrounding modern cockpit automation concerns.

Information Processing and Aircraft Automation

In the past, most ventures in automation design were technologically driven, with minimal consideration given to the human element of the interface. Recently, however, the need to examine this latter factor has become increasingly evident. Essentially, there is a relatively new and quite vital interest in human information processing in the presence of cockpit automation.

The fundamental information processing role of the human operator in the cockpit is to make decisions upon which to act. The automated aviation environment is unique, both in its decision requirements and in its decision consequences. It offers researchers the opportunity to examine a variety of information processing behaviors as they interact within an environment where both speed and load stresses surrounding a critical decision are often quite high (such as take-offs, landings and system failures) (Hart and Bartolussi, 1984). In order to examine the potential causes and possible design and/or training solutions to automation-related performance failures, it must discern how the operator, under time pressure, gathers information, how he/she interacts with the automated portion of the system, and how he/she reaches a final decision on a course of action; a decision which will lead to either success or failure.

As important as these inquiries may be to understanding the task environment of the aviator, until recently, there have been few attempts to venture into this unique cognitive realm. However, with increased interest in developing decision aids and decision programs in aviation, more and more investigators have acknowledged the need "to build an information base concerning the cognitive and information-processing aspects of flight" (Braune and Trollip, 1982).

Human decision making behavior is a complex, context-dependent arena, but there is a consensus of opinion in the literature that the process generally proceeds through a characteristic set of stages. Flathers, Giffin, and Rockwell (1982) have outlined four stages of decision-making to include "detection," diagnosis," "decision," and "execution." Other authors have proposed similar stages, often in the context of modeling human information processing within complex systems. Wickens and Flach (1988), for example, include within their model of information processing the sequence: "perception and attention," "situation assessment (diagnosis)," "choice," and "action." A similar concept has even been incorporated into a "real time expert advisory system" (Barrett and Donnell, 1990, p. 15) known as KOALAS, or "Knowledgeable Observation Analysis-Linked Advisory System" (Barrett and Donnell, 1989; Harris, Owens, Barrett, Parisi, and Becker, 1991). Within the "KOALAS taxonomy of intelligent control processes" (Harris et al., 1991) are the stages "sense," "interpret," "decide," and "act." One can therefore see the obvious importance of considering the separate and distinct elements of the overall decision-making/problem-solving process.

The literature is also consistent in its depiction of underlying cognitive structures affecting each stage in the decision process. Such structures are referred to by various terms: "internal models" (Braune and Trollip, 1982), "schemata" (Rumelhart and Ortony, 1977), "frames" (Minsky, 1975), etc., but all serve to establish patterns of associated events that reduce an operator's search for information by directing attention away from irrelevant and/or redundant cues. Such a reduction is necessary, as humans are not very qualified to process large amounts of information simultaneously and make sense of it (aside from the perceptual realm, obviously). This is due largely to limitations of human working memory and attention, which are further limited under conditions of speed and/or cognitive load stressors.

Two areas of recent concern relating to human attention and information processing within the cockpit include the effects of aircraft automation upon mental workload and situational awareness.

Aircraft Automation and Mental Workload

The relationship between attention demanded by a task and attention available from the operator's resource pool is expressed as Mental Workload. Sanders and McCormick (1987) discuss three distinct classes of attention that have relevance to the design and study of automation implementation: selective attention, focused attention, and divided attention. Selective attention involves the monitoring of many sources for a specific event. Focused attention requires the operator to monitor relatively few information sources within the context of many, but without being distracted by sources irrelevant to the task at hand. Finally, divided attention involves "time-sharing" (p. 66), where the operator is required to perform multiple tasks concurrently.

As such, it is this time-sharing that is of particular interest in studying the aviation environment.

A popular information-processing model with respect to divided attention is that of Wickens (1984). This model adopts a multiple-resource approach to divided attention, in which the author discusses three categories of attention that he believes are involved in different types of tasks, and which effect one's ability to divide attention among those tasks. These three categories are labeled as Stages, Modalities, and Processing Codes, and include within them distinct dimensions of attention resources. Stages refers to resources as belonging to either a perceptual/central processing category or a response selection/execution category. Modalities refers to the sensory medium over which information is presented (i.e., auditory vs. visual). Processing Codes indicates differences between spatial and verbal memory coding.

It is expected that timesharing between tasks of the same dimension within the same category would be less efficient than timesharing between tasks of different dimensions, as in the former case, attention for both tasks would be demanded of the same, limited resource pool.

Multiple resource theories such as that just described help explain why operators of two or more concurrent tasks tend to perform less well on one of the tasks. According to Wickens (1984), while single resource theories, which suggest one general pool of limited attention capacity, explain such decrements in timesharing performance equally well, they fail to account for three common findings (Sanders and McCormick, 1987, p. 67):

(1) why tasks that require the same memory codes or processing modalities interfere more than tasks not sharing common codes and modalities;

(2) why with some combinations of tasks increasing the difficulty of one task has no effect on performance of the others;

(3) why some tasks can be timeshared perfectly.

Theories of limited attention capacity, be they from a single or multiple resource, imply that there are certain levels and types of subtasks within an operating environment that may exceed said resources, thereby increasing workload.

There are two very general categories of workload: physical and mental. In the semi-automated cockpit, it is the latter concept, mental workload, which is of particular interest to issues regarding the design of aircraft systems.

The importance of understanding mental workload in aviation is considered so important that the Air Force currently employs workload measures in system designs. Additionally, the Federal Aviation Administration is soon to demand that aircraft be certified with respect to the level of workload imposed in the cockpit (Wickens, 1984).

Why is the understanding and measuring of mental workload considered so critical? There are two reasons, both affected by aircraft automation; one involves issues of economics, the other involves issues of safety and human performance.

Aircraft Automation, Workload and Cost Effectiveness

Aircraft automation has undoubtedly decreased economic costs in several areas. For instance, automation is argued to reduce the workload demands of the cockpit environment. It then follows that fewer aircrew members should be needed to carry out the operations in an automated cockpit. Because costs for training and compensating an additional crew member can be quite high, the economic benefits of such a reduction are obvious. However, one must question whether the reduced aircrew members would actually be capable of managing the new distribution of workload. In the case of the airline industry, aircrew size is typically reduced by about one-third without increasing workload on the remaining crewmembers (Parasuraman et al., 1990). Kantowitz and Casper (1988) suggest that the omission of the third member, generally the flight engineer, is warranted as "adding a third crew member does not reduce workload by one-third, since the need for communication among crew members may be increased" (p. 159). Additionally, they indicate that moderate levels of total crew workload may lead to underload of the third member, a condition also susceptible to increased probability of human error.

This introduces the second relevant issue, that of safety and human performance.

Aircraft Automation, Workload and Safety/Human Performance

Mental workload has been established as a significant factor in human error:

> The human is most reliable under moderate levels of workload that do not change suddenly and unpredictably... When workload is excessive, errors arise from the inability of the human to cope with high information rates imposed by the environment. When workload is too low, the human is bored and may not attend properly to the task at hand, also leading to error. (Kantowitz and Casper, 1988, p. 159)

There are several estimates regarding human capacity to transmit or manipulate simultaneous information inputs. All estimates, however, tend to reflect a rather small amount. The human tends to transmit information at about the rate of 10 bits per second (Kantowitz and Casper, 1988), with one bit of information being equivalent to "the amount of information required to decide between two equally likely alternatives" (Sanders and McCormick, 1987, p. 45). In terms of amount of information that can be simultaneously manipulated, one usually turns to the classic "seven plus or minus two" capacity of working memory (Miller, 1956), as working memory, as its name implies, is that used to currently manage and process information inputs. Recent research with regard to the capacity of working memory, however, suggests that the "magical number seven" is not maintained in the context of dynamic tasks. Instead, much of the evidence indicates the capacity of dynamic working memory is limited to about three "chunks" of information (Neville and Moray, 1986).

One can therefore see why dynamic environments, such as in modern aviation, present a challenge to human operators in terms of mental workload, but they also present a challenge to researchers in their attempts to measure this workload.

The three most common forms of mental workload metrics are physiological measures, subjective ratings, and secondary task measures (Kantowitz and Casper, 1988). Physiological measures include a variety of indices, such as heart rate, evoked potentials, pupil diameter, etc. Subjective ratings involve recording the subjects' scaling of particular tasks and/or events. Finally, secondary task measures include a variety of indices which focus upon variations in the performance of a secondary task with changes in the workload imposed by the primary task.

Although each of these metrics has advantages and disadvantages that can be carefully outlined with respect to controlled laboratory experiments, it becomes more and more difficult to apply this limited class of metrics to complex, dynamic tasks, such as aviation.

Hart and Bartolussi (1984) were prompted to examine subjective workload assessments of pilots in different phases of flight because of their belief that the various subtasks involved in a typical flight cannot be compared on the same dimension, with respect to workload. "Flying, like other complex tasks, is composed of a hierarchy of subtasks, each of which imposes specific demands and requires different levels and types of effort from a pilot" (p. 546). Additionally, the authors indicate that the total workload of a complex task is not necessarily the predicted sum of its parts.

The investigation utilized twelve instrument rated, general aviation pilots in 24 flights, each of which was further divided into 12 mini-scenarios. The flights were presented in written form, via a computer terminal, through which the subjects also entered their subjective ratings in response to the described events. In addition to evaluating overall workload of different segments of a flight, the subjects were asked to estimate the effects of certain events occurring during a particular phase of flight, based on their experience. Some significant findings resulting from the study included differences in overall workload requirements for the various flight segments (with take-off and landing being rated as highest in imposed workload), as well as for specific events occurring within particular segments. Additionally, it was found that pilot errors were classified by the subjects as an important source of workload, where such errors are typically viewed in research only as a measure of workload. The authors, in their discussion, indicate that their study demonstrates the extreme complexity of the workload construct, particularly with respect to complex task environments.

The Hart and Bortolussi (1984) study is based, in part, upon information gained from a preceding investigation conducted by Hart, Childress, and Bortolussi (1981). These investigators found that pilots were able to classify workload in terms of cognitive, physical, and perceptual aspects, and that, in fact, the subjective ratings for various flight tasks were compared by the subjects to different standards. The authors report that the pilots seemed to compare the scenarios with relative mental images of normal or exemplary workloads associated with the particular task type.

Aircraft Automation and Situational Awareness

Along with potential benefits through decreased workload with automation, one must consider the possibility of potential decrements through decreased situational awareness. Like workload, the term "situational awareness," though used by many, can be difficult to define. However, because a detailed explanation is critical to understanding its role with respect to the current document, a well-delineated definition has been adopted here.

Situational Awareness (SA) is defined by Endsley (1990) as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p.5).

The fact the term "situational awareness" is so commonplace to some degree reflects its importance in the world of aviation. Adequate SA is considered by aircraft developers to be one of the most important factors in both safety and mission effectiveness. It follows that a loss in SA has been found to be a major contributor to incidents and accidents in both commercial and military aviation. It is particularly associated with combat losses in the latter (Freeman and Simmon, 1991; Shaw, 1985). As such, both military aviation officials and the FAA consider the problem of loss of SA one of major concern (Endsley, 1991). According to Endsley (1991) "even the best trained and most experienced pilots can make the wrong decisions if they have incomplete or inaccurate SA" (p. 2).

Because automation of a task relegates the human operator to the role of passive monitor, much potential exists for a loss of SA. Yet one must have adequate SA in order to maintain an accurate representation of the status of information. It is a fundamental component in effective cockpit decision-making.

Model of Stages Involved In A Decision Within A Complex, Semi-Automated Cockpit

The theoretical model of the stages involved in a decision within a complex, semi-automated cockpit is displayed in Figure 1. The pilot within such a system is exposed to a variety of information sources. Three sources that he/she directly monitors are the instrument panel (indirect sensations such as fuel level), the external environment (direct sensations such as engine sounds), and the internal environment (direct sensations such as physiological responses to increased body weight). The pilot also has an automated aid capable of handling partial monitoring and control functions. This aid attends to state of the aircraft information with respect to a predetermined "internal model," its program, and, in the event of an inconsistency, informs the human operator and/or performs direct control modifications.

Because of the amount of information present and the limited processing resources of the human operator, not all available information is attended to by the pilot during Stage 1. An internal model based primarily upon training and experience, as well as the continuous input of sensory information, and consisting of expectations, probabilities, intra- and inter-model relationships, feeds into the human processor and guides his/her attention through scanning behavior.

Barrett and Donnell (1989) define such an internal model as "a formal representation of a natural process" (p. 17). They indicate that certain aspects of natural process that must be represented within the internal model cannot be directly evidenced by sensors, human or automated, and must therefore be <u>deduced</u>. Such deduction, they argue, is the forte of automation, due to its ability to assimilate great amounts of information very quickly. However, the authors also propose that the formation of the model itself is through a process of <u>inductive</u> reasoning, which is argued to be more the forte of human operators.

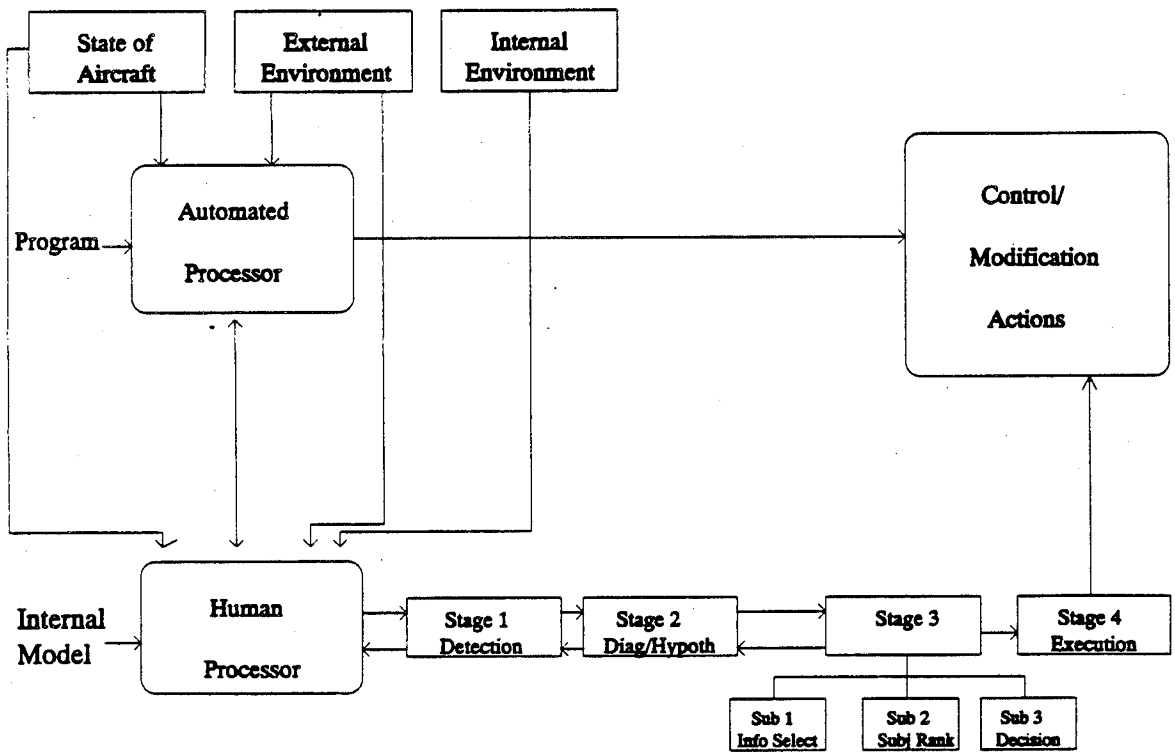


Figure 1: Theoretical Model of Stages Involved in Decision-Making within a Complex, Semi-Automated Cockpit

According to Harris, Owens, Barrett, Parisi, and Becker (1991):

> The most important issue in the design of human-mediated weapon control is the definition of the human operator's role in the sensing, interpretation, decisionmaking, and action processes of the control system being designed. Since sensing, decisionmaking, and action processes in the KOALAS taxonomy are defined to be deductive, these processes can be largely (or wholly) automated; it is in these areas that machine intelligence offers greatest payoff in the control of weapons systems. The crucial human role in the system is in the interpretation process, a function that can be aided, but not automated. (p. 7)

Upon detection of a system fault, the pilot moves into Stage 2, the diagnostic/hypothesis-testing stage of decision making. During Stage 2, the pilot searches for information to explain that which does not fit in the current internal model or frame. In doing this, the operator searches for plausible models to be applied to the situation and to redefine the pattern of information to which the pilot must readapt his/her scan in order to successfully solve the encountered discrepancy. This may ultimately involve a model modification or transformation, in which, respectively, the current model is adapted to account for the new data, or a new, more appropriate model is selected to guide resultant information sampling and decision making behavior.

"The matching process that decides whether a proposed frame is suitable is controlled partly by one's current goals and partly by information attached to the frame... A frame, once evoked on the basis of partial evidence or expectation, would first direct a test to confirm its own appropriateness" (Minsky, 1975 , p. 159).

Once the situation-appropriate frame or model is selected and/or adapted, the pilot is then able to move to Stage 3, the functional decision process. Stage 3 involves three substages. The first substage, information selection, is a direct result of the resolution of Stage 2; the selection of the new model determines the relevant information to be attended. Substage 2 involves subjective rankings of this information in terms of the final decision. Certain sources of information will be considered more vital by direction of the newly evoked model. This is dictated by expected reliabilities of information and probabilities of outcomes associated with the information constraints and goals of the model. Substage 3, or the functional decision, involves the actual choice of what action to undertake. Stage 4 is the execution of the control or modification action.

This study investigates the effects of automated cockpit systems on human performance as a function of the stability of the internal model the human employs to operate a task. The focus of this examination is a bit different from that of other studies within the cockpit automation domain. The purpose of this study is to examine the differential effects of removing, via automation, the human operator "from the loop" and requiring him/her to suddenly reenter that "loop" within the confines of the particular subtasks involved. The relevant characteristics of the subtasks do not only concern the general mental demands imposed. Instead, the present investigation asks whether the more important demands are those dealing with the crucial decision stage of the subtask and the dynamic versus static manner in which task information is updated within the context environment.

Statement of the Problem

Internal models by whatever term they are referred, are composed, most fundamentally, of learned associative relationships. In automated systems, the removal of the human operator from the "loop," so to speak, reduces the opportunities that are necessary to continually update the internal model, resulting in a lack of direct situational awareness for the particular task. Barrett and Donnell (1989) refer to updating of the internal model as "model transformation," and suggest that "in military systems this process is referred to as the process of maintaining situational assessment" (p. 17). Without such constant situational assessment or awareness, the human operator would be dependent upon the "internal model" provided by the automated component. As humans are, once again, poor passive monitors (Chambers and Nagel, 1985), this would lead to a situation where, upon reentering the loop, the human's internal model would not be "up to date," particularly if such reentry was unexpected or sudden, as in automation failure. Such a situation would have an effect, most directly, on the detection and diagnosis stages of decision making, as one must detect the need to reenter the loop, and then diagnose the problem if one exists. Although the situation of returning to manual performance following automation can be expected to have an effect on the detection and diagnosis stages of the decision-making process, it would be expected to affect each stage differently.

It is proposed that decision tasks directed by a more stable internal model are more dependent on the detection stage of the decision-making process, whereas those decision tasks directed by a more dynamic internal model are more dependent on the diagnosis stage. Recall that in the case of a stable internal model, the information relevant to decision-making, once learned, does not change across time. Although humans involved in tasks with stable internal models may poorly monitor the automated signals provided, upon reentering the loop, the model he/she had when automation began will still be relevant, as the information has not changed across time. For example, imagine an operator monitoring the automation of a simple system monitoring task. The task is one which the operator has often manually performed, and the rules for what a signal is and how to respond to that signal, once learned by the operator, do not change. During automation, the signals are responded to by the automation, and the human operator only monitors the operability of the automation. In the case of automation failure, the human would only have to detect the failure and realize he/she must begin manually responding to signals. Once this was detected, the operator could immediately call upon the internal model which he/she had used in the past to operate the task, as the model would not have changed during the time the automation was controlling the task. Therefore, the crucial stage in reentering the control loop is detection of a fault or problem, and/or the fact that one must regain manual control. With a stable internal model, information used in the diagnosis stage to further direct decision actions remains current, and this stage is essentially bypassed.

In the case of a dynamic internal model, the information relevant to decision-making, once learned, does change across time. The information cues and control responses in that task carry a history component. The decision factors are not stable, but depend upon past events and the current state of the overall system. For example, in controlling a fuel management system, the operator must consider the desired fuel state within a particular fuel tank and the various paths to achieve that state. The path the operator chooses will be dependent upon many factors such as the fuel state of auxiliary tanks. In order to maintain balance within the system, the operator must understand the dynamic relationship between the fuel sources within the system, and this relationship changes across time and events. When the human responsible for a task with an associated dynamic internal model poorly monitors the automation performing said task, he/she does not continue to update the information which is changing. Therefore, upon reentering the loop, he/she is left with an "out-dated," irrelevant model which must be updated or modified. This updating takes place in the diagnosis stage. The human operator must first detect the problem, then diagnose the problem. This stage cannot be effectively bypassed if one loses situational awareness of a task with an associated dynamic internal model.

Performance on a task with an associated stable internal model and dependent on the detection stage, should be greatly affected by the degree of interaction of the human operator with the task. This can be predicted by classic vigilance work, where a predominant finding states that detection performance is enhanced with increased certainty of signal occurrence (Smith, Warm and Alluisi, 1966; Warm and Alluisi, 1971). One way of increasing such certainty is to actually increase the frequency of signals per unit time, so that the internal model used to direct decision making may be continually updated. However, if the human operator is removed from the loop such that he/she is monitoring the automated monitor, rather than the signal task itself, then the internal model used to direct the decision making is dependent upon the probabilities associated with the reliability of the automated system, in addition to the actual probabilities of signal occurrence. One would therefore expect that in the case of manual monitoring, the predictions based on classic vigilance research would be upheld in terms of human responses. That is, the human operator's ability to monitor would be primarily affected by aspects of time-on-task and the signal properties (signal rate, background event rate, signal salience, etc.) (Parasuraman and Davies, 1977). In the case of the automated condition, on the other hand, human detection of a signal (the properties of the signal now being such that they reflect reliability of automation) would be affected by the operator's internal model of the automated system as well. In other words, his/her expectations of automation performance and reliability would affect which cues he/she is directed to attend. In the case of full automation failure, the signal properties of the task, as well as the expectations of the task, would be expected to affect human operator performance in a manner different from that expected of a human operator who has been manually monitoring a task. Performance would be expected to deteriorate somewhat following automation failure of a stable task (due to prior reliance on the automated system) but the engagement level (degree of interaction required between operator and task) associated with the signal rate of the task should remain the same. In other words, although the higher engagement levels should still produce better performance following automation failure (as in full manual conditions), overall performance should be somewhat lessened following automation failure. However, significant performance detriments would be expected only in the case of low engagement levels.

In contrast to the detection stage of decision-making, the diagnosis stage would be more affected by the stability of the internal model that directs the decision making associated with the task. The more dynamic a task, the more changeable the associated internal model. As a result, the relationships between decision-relevant sources of information about the task consistently change. One would therefore expect that the cognitive load on a human operator who has to reenter the loop suddenly and unexpectedly, as in an automation failure, would be greater since he/she would be forced to consult an internal model that first must be updated to establish current situational awareness. In the case of a less dynamic task, (one in which relationships and properties within the task, once learned, do not change) such reentry would be easier. The internal model directing situational awareness would be more stable, requiring little if any modification. Additionally, one would expect to see, with a dynamic task, the opposite effect of engagement level (degree of interaction between operator and task) on human performance following automation failure than would be seen in a task with a more stable model. In the case of a task with a more dynamic internal model, the level of engagement would affect two important factors: task demand and task involvement. In the case of full manual control within a multitask situation, one would expect task involvement to have more effect upon performance. The human operator would benefit from a higher engagement level due to greater interaction (and thereby more frequent updates) with the internal model (enhanced situational awareness). By contrast, following automation of the same dynamic model task, one would expect that the sudden change in automation status of the task (and the subsequent requirement for the operator to unexpectedly perform all tasks manually) would establish a situation in which the task demand factor of the engagement level would reach primary importance. Therefore, one would expect that higher engagement levels (i.e., more interaction required between operator and task) in the dynamic model task would decrease performance following automation failure (due to increased cognitive load).

Goals of Investigation

1. Examine the effects of automated tasks and general automation failure on performance of non-automated tasks.

2. Examine the effects of automation failure on subsequent manual performance of particular subtasks as a function of the crucial decision stage (based on the associated stability/dynamism of the internal model) of those subtasks.

3. Examine the effects of the degree of interaction required between operator and subtask (Level of Engagement) on performance of that task. Also, examine the effects of automation failure of a particular subtask on subsequent manual performance of that subtask as a function of the subtask level of engagement.

4. Examine the effects of the task type and engagement level (Condition), judgment type, automation, and automation failure of various subtasks on situational awareness and workload.

5. Examine the effects of Judgment Type (simultaneous or **absolute** versus successive or **comparative** judgment) on performance and to examine the effects of automation failure on subsequent manual performance of particular subtasks as a function of whether that subtask is of a comparative or an absolute judgment type.

6. Examine the effects of time-on-task (Block Number) on performance of that task.

These goals were addressed in two separate experiments. The first examined performance data following automation failure (with respect to full-manual and part-manual controls). The second experiment examined situational awareness issues using the Situational Awareness Global Assessment Technique (SAGAT).

Statement of Hypotheses: Experiment I

The specific predictions of this experiment were based upon five primary hypotheses: a main effect for task type, an interaction between task type and engagement level, a main effect for judgment type, an interaction between judgment type and condition (consisting of task type and engagement level combination), and a main effect for block number.

These predictions are explained with respect to the three tasks used in the investigation. These three tasks included a system monitoring task, a resource (fuel) management task, and a tracking task. These tasks are described in detail in the Methodology.

The system monitoring task was manipulated in the investigation because it was associated with a stable internal model. The resource (fuel) management task was manipulated in the investigation because it was associated with a dynamic internal model. The tracking task was not manipulated in the investigation. It was always performed manually by the subjects, and its characteristics never changed.

1. A main effect for task type was predicted:

(a) Automation of either the system monitoring or the resource (fuel) management task was predicted to improve performance on the non-automated tasks. However, automation of the resource (fuel) management task was expected to have a greater positive effect on performance of the non-automated tasks. This was predicted because it was hypothesized that a dynamic internal model (associated with the resource [fuel] management task) is more dependent upon the diagnosis stage of decision making, whereas a stable internal model (associated with the system monitoring task) is more dependent upon the detection stage. As the diagnosis stage involves a deeper level of decision processing, operation of the resource (fuel) management task was expected to produce a greater overall cognitive load. Automating the resource (fuel) management task should therefore allow greater concentration on the non-automated tasks.

(b) With respect to workload, it was expected that overall workload would be highest when all tasks were performed manually for the entire session (the full manual control conditions). This would be followed by the condition in which the system monitoring task was automated. Finally, the lowest workload was expected from the condition in which the resource (fuel) management task was automated. The resource (fuel) management task was expected to contribute more to overall workload than the system monitoring task due to its dependence on the deeper level of the decision-making process: diagnosis, as well as the requirement to continually update the internal model associated while engaged in the resource (fuel) management task.

2. An interaction between task type and engagement level was predicted:

(a) It was predicted that automating the system monitoring task (a stable internal model task) would be detrimental to a subject's ability to regain manual control of that task following automation failure only when that task was of a low engagement level (lower signal rate). This was predicted because, although automation of the system monitoring was expected to reduce the subject's situational awareness of the task while automated, it was also expected that, once the subject had detected automation failure, he/she would be able to immediately begin manual operation of the task on the basis of the internal model that was present before automation of the task. In other words, because the internal model of the system monitoring task is stable, and the meaning of and responses to information does not change over time, the subject should only have to detect the need to regain manual control. No internal model modifications would be necessary. Performance subsequent to automation failure would then be expected to be dependent on factors related to detection, such as signal rate. As such, classic vigilance work would lead to the expectation of lower performance following low engagement levels (low signal rates) in the system monitoring task (Parasuraman and Davies, 1977). It was also expected that performance on the two manual tasks (resource management and tracking) would be unaffected following automation failure of the system monitoring task.

(b) Automating the resource (fuel) management task (dynamic internal model task), on the other hand, was expected to have detrimental effects on the subject's ability to regain manual control of that task under both low and high engagement levels (fuel pump failure rate) of the task. This was predicted because automating the resource (fuel) management task was expected to reduce the subject's situational awareness of the task. Unlike the stable internal model task, while the subject is not maintaining situational awareness of the dynamic internal model task, the meaning of and responses to the task information are changing. Therefore, when automation fails, the subject has not only to detect that he/she must regain manual control, but he/she must also diagnose the present state of the system to decide upon the best way to intervene. Updating the internal model guiding this decision process is first necessary. Furthermore, it was predicted that when the resource (fuel) management task was automated, performance on the two manual tasks (system monitoring and tracking) would also decline following automation failure. It was expected that the cognitive demand resulting from the need to update the resource (fuel) management internal model would deter from subject performance on the remaining tasks.

(c) Engagement level was predicted to affect system monitoring in a consistent manner. Specifically, performance on the system monitoring task was expected to be worse under low levels of engagement (again, due to findings from classic vigilance studies).

With respect to the resource (fuel) management task, on the other hand, the engagement level was predicted to affect performance in an inconsistent manner. Specifically, it was expected that in blocks 1-4, under conditions where either system monitoring or resource (fuel) management was automated (except in the controls), a high level of engagement in the resource (fuel) management task would benefit performance on that task. This was predicted because the higher level of engagement (higher number fuel pump failures) would force more interaction between the subject and the resource (fuel) management task, thus allowing for more frequent updates of the dynamic internal model associated with the task. By contrast, it was expected that a high level of engagement in the resource (fuel) management task would be detrimental to regaining manual control of that task following its automation failure. Performance under the condition where resource (fuel) management had a high engagement level and had been automated during blocks 1-4 should deteriorate in block 5, and perhaps into block 6. Again, this is unlike the system monitoring task. This discrepancy is because in the case of the system monitoring task, it was hypothesized that the crucial decision stage for subjects would be detection. Once the automation failure of the system monitoring task was detected, subjects would be able to call upon their previously used internal cognitive model, and it would still be accurate, due to its stability. Therefore, following automation failure of the system monitoring task, the higher engagement level of that task should still enhance performance. On the other hand, in the case of the resource (fuel) management task, the subjects would not only have to detect that resource (fuel) management automation had failed and they would need to regain manual control, they would also have to diagnose the situation and make modifications to their out-of-date internal model. Therefore, following automation failure of the resource (fuel) management task, the higher engagement level of that task should only add to the already high cognitive load imposed by the need to regain an understanding of the situation and to modify the internal model in the time-stressed, multi-task scenario.

(d) With respect to workload, engagement levels of the tasks in both the full manual control conditions and the automation conditions were expected to have the following effect, in order of decreasing workload: high engagement level resource (fuel) management task with low engagement level system monitoring task, high engagement level resource (fuel) management task with high engagement level system monitoring task, low engagement level resource (fuel) management task with high engagement level system monitoring task.

3. A main effect for judgment type was predicted:

(a) With respect to performance, comparative judgment types were predicted to enhance performance over absolute judgment types in both the system monitoring and the resource (fuel) management tasks.

b) With respect to workload, absolute judgment types were expected to impose a higher overall workload across conditions.

4. An interaction was predicted between judgment type and condition (consisting of task type and engagement level) for performance only:

The detriment to performance imposed by absolute versus comparative judgment was expected to be more pronounced in the system monitoring task. This was predicted on the basis that the absolute judgment type would add a degree of instability to the stable model associated with the system monitoring task (from possible inconsistencies due to inability to maintain absolute judgment criteria accurately in working memory). Therefore, the quality of the overall model would be changed to one less stable (more dynamic).

In the case of the resource (fuel) management task, on the other hand, with the absolute judgment type a degree of instability would be added to an already unstable (dynamic) internal model. The detrimental effects were expected to be less extensive, because only the degree, not the actual quality, of the model would be changed.

5. A main effect for block number was predicted:

The subjects performed each condition for thirty minute sessions. Each session was divided into six, 5 minute blocks. During blocks 1-4, either the system monitoring or the resource (fuel) management task was automated, except in the full manual control conditions, where all three tasks (system monitoring, resource (fuel) management and tracking) were performed manually by the subject for the entire 30 minute session. Blocks 5 and 6 consisted of the two 5 minute periods following automation failure (except, again, in the case of the full manual control conditions, in which case blocks 5 and 6 simply represented continued time-on-task). Performance was expected to be generally better in earlier blocks, due to a loss in vigilance over time-on-task.

Statement of Hypotheses: Experiment II

The specific hypotheses for Experiment II are based on Endsley's (1990) definition of situational awareness; in particular, levels I and II. Level I, again, refers to the subject's perception of events in a task. Level II refers to the subject's understanding of the meaning of events in a task.

The following specific hypotheses were generated:

1. All percent correct scores for Level I (Perception) SA questions about both the resource (fuel) management and the system monitoring tasks were expected to be lower under the absolute judgment type, relative to the comparative judgment type. This was predicted because, in a multi-task situation such as the MAT battery, it is expected that subjects will have more difficulty perceiving relevant information (i.e.,

signals and out-of-normal range fuel levels) in the absolute judgment type, where there is no visual referent defining the normal range. In other words, when the desired range is memory-dependent.

2. It was predicted that during automation of the system monitoring task, there would be a loss of Level I (Perception) SA. This prediction stems from the hypothesis that the system monitoring task is associated with a stable internal model. Furthermore, this discourse argues that the most crucial stage in the decision-making process guided by a stable internal model is detection. Because the simulations in Experiment II were run for 17 to 27 minutes before subjects were queried on the SA questions, one would expect to see a vigilance decrement, particularly in the automated task, where the human is serving as supervisory monitor. Therefore, one would expect subjects to answer fewer questions regarding perception of events in the system monitoring task correctly while the task was automated.

3. It was predicted that during automation of the system monitoring task, there would be a loss of Level II (Meaning) SA **only** under the absolute judgment type. Again, this follows from the hypothesis that the system monitoring task is associated with a stable internal model and as such, the meaning of the decision-relevant information, once learned, does not change across time. However, recall that it is argued that, in the case of the absolute judgment type, a degree of instability is added to the model, because the decision-relevant information is memory-dependent. Therefore, although the meaning is stable across time, the human memory is not as stable. If there is a potential for change in the meaning of information over time,

whether this change is actual or due to the failings of memory, then there is an increased likelihood of losing Level II (Meaning) SA following long periods of removal from direct control of a task.

4. It was predicted that during automation of the resource (fuel) management task, there would be a deeper loss of SA; specifically, Level II (Meaning). This prediction stems from the hypothesis that the resource (fuel) management task is associated with a dynamic internal model. It is further argued that the crucial stage in the decision-making process associated with a dynamic internal model is diagnosis. Diagnosis is a deeper stage in decision-making than detection. Therefore, even if subjects remained somewhat vigilant of the automated resource (fuel) management task (i.e., vigilant enough to detect problems), they would need a deeper level of SA, specifically, an understanding of the meaning of information and how it has changed over time, in order to make an adequate assessment of the situation at the time they are queried. It is argued that because an understanding of the meaning of information in the resource (fuel) management task is not needed while the human is serving as passive monitor (and not a very good passive monitor, at that), Level II SA is more easily lost during the period when the resource (fuel) management task (or any such task with a dynamic internal model) is automated.

CHAPTER II

METHODOLOGY

Subjects

Thirty-two volunteers from the Naval Air Warfare Center, Aircraft Division, Warminster served as subjects. The subjects possessed little to no experience piloting an aircraft, had 20/20 or corrected-to-20/20 visual acuity, and full color vision (one subject in first study reported a slight red-green color deficiency, but indicated that it did not affect his ability to read the screen). Of the sixteen subjects in the first study, all were male. Their ages ranged from 23 to 55 years, with a mean of 35 and a median of 32. Of the sixteen subjects in the second study, all but three were male. Their ages ranged from 21 to 50 years, with a mean of 31 and a median of 30.

Materials and Apparatus

Subject Briefing Packages

Prior to the first day of scheduled running for a particular subject, that subject was given a "Subject Briefing Package" and instructed to read and study the information prior to the first training session. The packages for Experiments I and II are included in Appendices B and C, respectively.

Subjective Measures

Workload Measure

The NASA Task Load Index (TLX) is a subjective workload measure developed by the Human Performance Research Group at NASA Ames Research Center (Hart and Staveland, 1988). Research has shown the NASA TLX to be a reliable and sensitive measure of overall workload (Vidulich, 1989; Vidulich and Tsang, 1987).

When using the NASA-TLX, subjects rate workload levels for six different subscales: Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration (Appendix D includes the definitions of the subscales). Each subscale is rated from "Low" to "Hi" (except for own performance, which is rated from "Good" to "Poor"), and the several tick marks on each subscale each represent 5 points, for an overall 100 point range. An overall workload rating is computed based on a weighted average of the ratings on these subscales.

The NASA-TLX was presented via an IBM-compatible 386 personal computer. Figure 2 displays a sample NASA-TLX. Each tick mark represents an increment of 5 in a series of 100. This was explained to each subject. Each subject placed his/her inputs by using the arrow keys on the keyboard. For each subscale, the subject began with the pointer in the middle, neutral area of the scale. The left arrow key was used to move the pointer left on the scale; the right arrow key was used to move the pointer right on the scale. The up arrow key was used to stop movement of the pointer in either direction on the desired tick mark. Finally, the down arrow key was used to move the pointer between subscales (unidirectional).

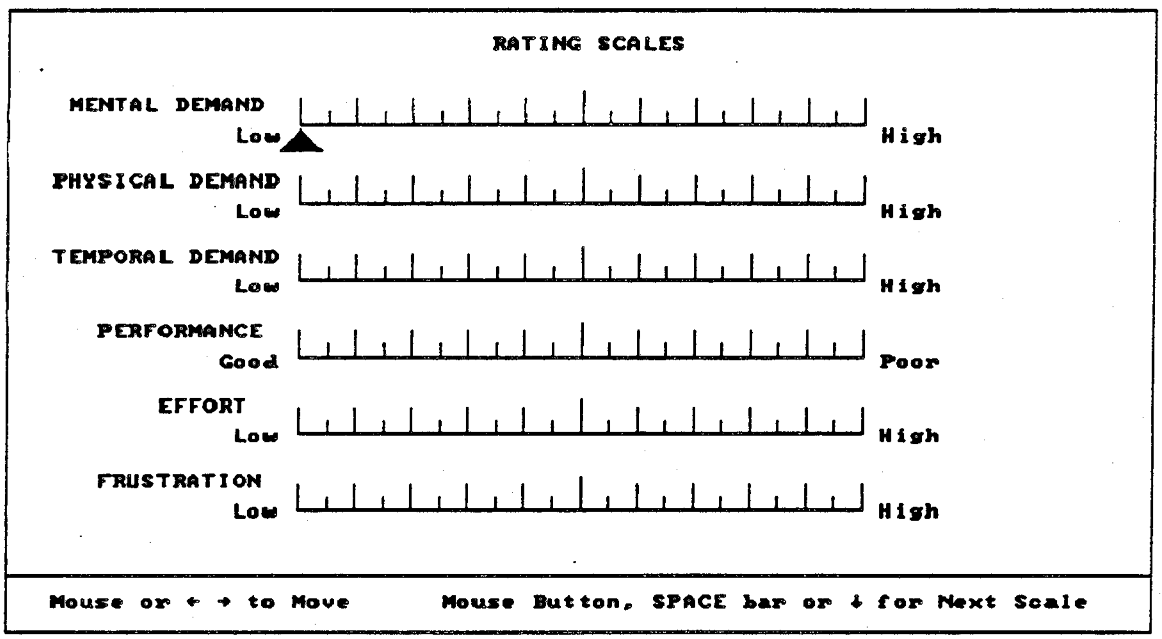


Figure 2: Sample NASA-TLX

All of this information was described to the subjects in their initial training session. Subjects were also provided with a sheet of paper containing definitions of all six subscales to which they could refer at any time. The NASA-TLX was given following each condition in the first experiment only.

Situational Awareness Measure

The Situation Awareness Global Assessment Technique (SAGAT) was developed in 1987 by Dr. Mica Endsley at Northrop in order to provide a means of collecting objective situational awareness data for man-in-the-loop tactical air simulations. The procedure involves stopping the simulation at some random point in time, blanking the screen, and asking the pilot (subject) a series of questions regarding the information present on the screen at the time the simulation stopped. According to Endsley (1990), "the comparison of the real and perceived situation provides an objective measure of pilot SA" (p. 12). Several stops may occur at random in order to gain the numbers of measures necessary for statistical significance.

As the original version of SAGAT was developed for tac-air simulations, a modified version was adopted for the present experiment. Six sets of questions were developed from random selection of a question pool (Appendix E) in order to query subjects.

Each set of questions was designed to examine hypothesized differences in loss of situational awareness (SA) between the system monitoring and resource (fuel) management tasks. Recall that the adopted definition of situational awareness defines three levels: loss of perception of events (Level I), loss of meaning of events (Level

II), and loss of future projection (Level III) (Endsley, 1990). Because the particular tasks were not very well suited to questions pertaining to future projection, most of the questions focused upon the division between Levels I and II. Each set of questions included two System Monitoring "perception" questions, four System Monitoring "meaning" questions, two Resource (Fuel) Management "perception" questions, four Resource (Fuel) Management "meaning" questions, one "future projection" question for each task, and one question pertaining to tracking (in order that subjects did not come to view this task as irrelevant). The order of questions was randomized across the six sets. One set of questions (chosen randomly) was given to each subject via pencil-and-paper following each condition in the second experiment only.

Performance Measures

Multiattribute Task Battery

The investigation employed an adaptation (Parasuraman, Bahri, and Molloy, 1991) of the Multi-Attribute Task Battery (Figure 3), or MAT (Comstock and Arnegard, 1990), which is a multi-task flight simulation package. The battery was presented via a 386, 33mhz 80386 personal computer.

Task Descriptions

The MAT task battery used in this investigation accesses three different, aviation-relevant, general information processing areas: perceptual-cognitive (via a system monitoring task), cognitive-strategic (via a resource [fuel] management task), and perceptual-motor (via a compensatory tracking task) (Parasuraman et al., 1991). As the present investigation was concerned with the effects of automation upon different cognitive processing levels, only the first two tasks were manipulated. Figure 3 displays the battery as it appeared to the subjects.

A further aspect of the MAT task battery that adds to its realism is the fact that the tasks are displayed dynamically. That is to say, rather than the events being presented discretely, the displays are updated continuously (Parasuraman et al., 1991).

Additionally, the presentation of the tasks is easily manipulated, as various "scripts" can be chosen or modified in order to vary subtask events in terms of sequence, type, and amount, so as to consequently vary workload. The modifications of Parasuraman et al. (1991) allow for further manipulations in terms of task automation. Their scripts allow the experimenter to tailor which tasks are automated, when, and for how long, as well as the reliability of the automation.

System Monitoring Task

As shown in Figure 3, the system monitoring task was presented in the upper left window of the terminal screen. The overall task consisted of two parts: warning light monitoring and probability monitoring. The detection of the lights was <u>not</u> used in this investigation. The green light always remained on, and the subjects were instructed to ignore the two lights. The second system monitoring task involved probability monitoring. As indicated in greater detail in Appendix B "Instructions to Subjects," each subject <u>was</u> instructed to perform this part of the task. Referring again to Figure 3, there were four vertical scales with pointers (colored yellow on the terminal screen) which moved vertically, indicating temperature and pressure of two

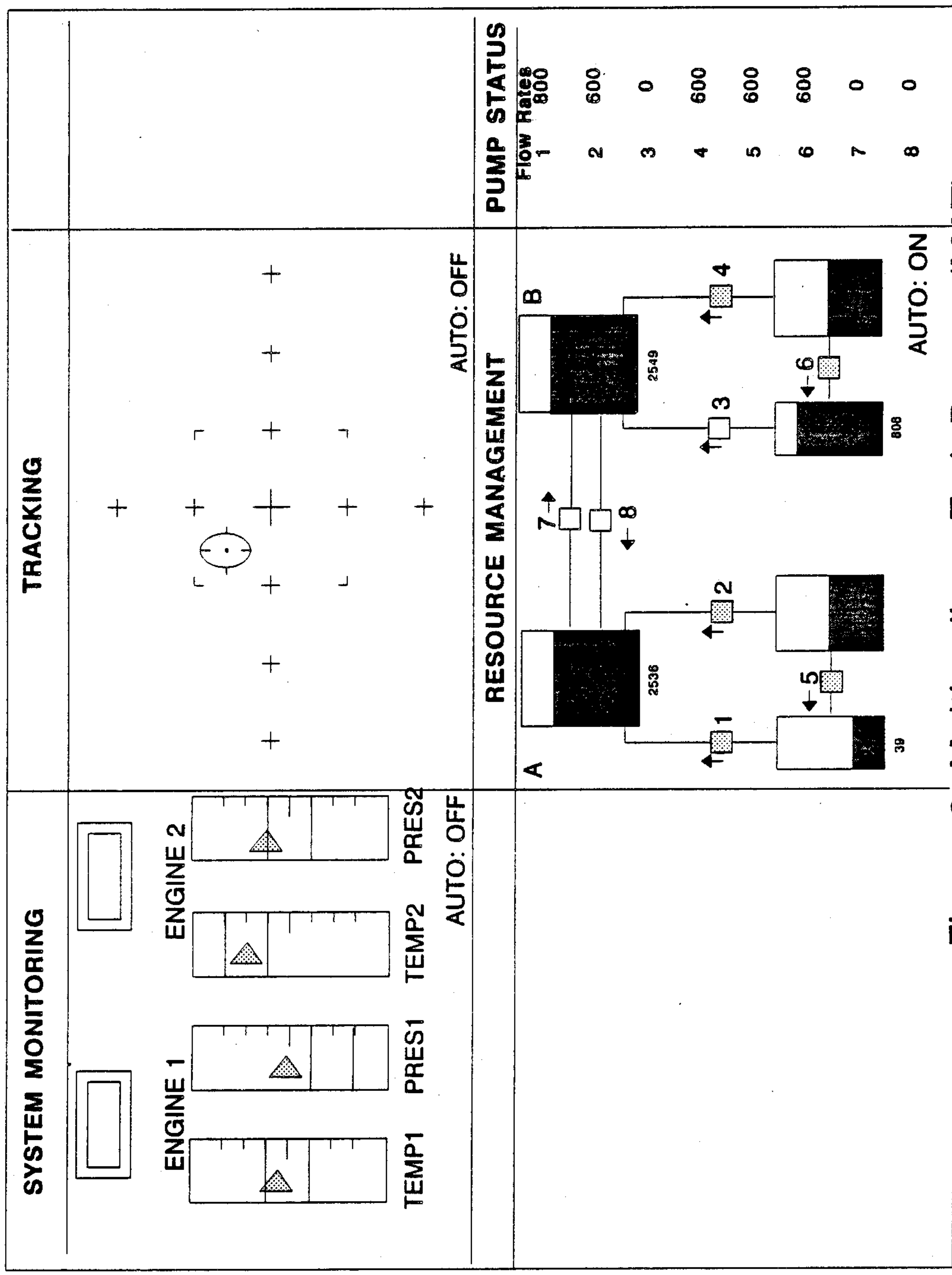


Figure 3: Multiattribute Task Battery (MAT)

aircraft engines. Under normal conditions, the pointers oscillated minutely around the normal ranges of their respective scales (within a 3-tick mark range). In the comparative judgment types, this area was demarcated by red lines which crossed the scale above and below the "normal" area. In the absolute judgment types, on the other hand, there were no such red lines.

At randomly scripted points in time, each scale's pointer shifted (independently of the others) to produce a "signal." A signal was defined to a subject as "any time the yellow pointer goes either completely above or completely below "normal" range." "Normal Range," in turn was dependent upon the scale. It was defined as "the middle three tick marks" for the first and fourth scales, "the lower three tick marks" for the second scale, and the "upper three tick marks" for the third scale. The normal ranges were staggered in this manner for the present investigation to prevent subjects in the absolute conditions from using adjacent scale pointers for comparison. Each subject was instructed to respond to signals by depressing the function key associated with the particular scale (the "F1" key for scale 1, the "F2" key for scale 2, the "F3" key for scale 3, or the "F4" key for scale 4). If a correct detection and response was made by the subject, the particular pointer returned to center position and did not oscillate for 1.5 seconds. If the subject did not detect the signal or responded incorrectly, the system returned itself to normal after 10 seconds out-of-bounds.

Under the automation mode, an "AUTO ON" sign was lighted in the lower right corner of the system monitoring window. When the system monitoring task was automated, signals still appeared at the same rate as in the manual version. However, the signals were detected and correctly responded to automatically, within 4 seconds. The subject was instructed there may be occasional instances of "suboptimal" automation, as well as the fact that he/she might be asked to report such failures in automation at a later time. Suboptimal automation was defined as cases in which it would take the automation slightly longer (in actuality, 10 seconds) to respond to a signal. There were 3 such instances of suboptimal automation within each condition of automated system monitoring. For the first fifteen minutes of each such condition, there was one instance of suboptimal automation per five minute block. The last such instance occurred at least 5 minutes prior to automation failure. Suboptimal automation was distinguished from total automation failure in that in the latter case, the "AUTO ON" sign in the lower right corner of the window would be extinguished, and signals would no longer be responded to automatically. Each subject was instructed that in the case of automation failure, he/she must regain manual control.

The number of signals in both the manual and automated conditions occurred according to the level of engagement in the experimental condition. In the high engagement condition, there were 12 signals per 10 minute block (6 per 5-minute block), occurring at random intervals between 30 and 90 seconds. In the low engagement condition, there were 6 failures per 10 minute block (3 per 5-minute block), occurring at random intervals between 60 and 120 seconds. System Monitoring signals were also separated from faults occurring in the Resource (Fuel) Management Task by a period of at least 10 seconds.

Performance measures for the system monitoring task included the following:

Percentage of signals correctly identified, number of false alarms, and reaction time were the three measures collected for the System Monitoring Task. These measures were then averaged in 5-minute blocks over the entire 30-minute session. Means were used for percent correct and false alarm measures. Median was used for reaction time measures. This latter measure of central tendency was chosen for the reaction time measure due to its relative insensitivity to extreme measures. Under conditions in which the System Monitoring Task was under manual control, all 6 blocks of measures (for that task) were used in the analysis. On the other hand, Under conditions in which the System Monitoring Task was automated, only the final 2 blocks of measures (for that task) were meaningful to the analysis (as only these blocks reflected subject performance). However, the effects of automating System Monitoring on the remaining two tasks (resource [fuel] management and tracking) were still examined across all 6 blocks of measures for those tasks.

Resource (Fuel) Management

As depicted in Figure 3, the resource (fuel) management task was displayed in the lower two right windows of the terminal screen. The larger of the two windows displayed six "tanks." The amount of fuel present within each tank was indicated both graphically and numerically. Graphically, a green "shaded" level increased or decreased, reflecting fuel changes. The amount of fuel was also indicated numerically, in gallons, beneath tanks A, B, C, and D, and this number was updated every two seconds. The unlabeled tanks had unlimited fuel capacity, and thus no changes in fuel level were indicated either graphically or numerically. Tanks A and B had a maximum capacity of 4000 gallons each, while tanks C and D had a maximum capacity of 2000 gallons each. Again, the remaining two tanks had no limit on capacity. The diagram in Figure 3 represents the manner in which the tanks were connected by pumps. Fuel could travel only in the direction indicated by the arrows next to the pumps. The task of each subject is to maintain the level of fuel in tanks A and B at a critical level of 2200 and 2850 gallons, respectively. As in the system monitoring task, these critical levels were staggered to support the absolute versus comparative manipulation, since the two main tanks were adjacent to each other. In the comparative judgment types, red lines marked the 2000 - 4000 fuel level on Tank A, and the 2650 - 3050, in order to provide a reference by which to make a comparative judgment. Each subject was instructed that the optimal level at which to maintain the fuel rested midway between the red markers, at 2200 and 2850 gallons in Tanks A and B, respectively. In the absolute judgment types, no lines of any color marked the desired fuel levels. Each subject was instructed to remember the respective levels to maintain. The rate of fuel flow out of Tanks A and B was 800 gallons per minute. Each subject was therefore required to adopt a strategy of transferring fuel from the remaining tanks in order to maintain the critical levels. This transfer was carried forth by activating (depressing the number key on the keyboard which corresponded to the number of the pump) the lower pumps. Again, each pump could transfer fuel in only one direction. While a pump was actively transferring fuel, it was colored green, whereas it was colored black when no fuel was being transferred.

The subject began each task scenario with 2200 gallons of fuel in Tank A, 2850 gallons in Tank B and about 1000 gallons in each of Tanks C and D.

The "Pump Status" window to the right of the resource (fuel) management diagram window indicated the flow rate of each of the pumps. There were eight pumps, listed numerically in the first column. Adjacent to each pump number was its flow rate. A rate of zero was indicated when the pump was off, otherwise, the rate was dependent upon the pump number. Pumps 1 and 3 each had a flow rate of 800 gallons per minute. Pumps 2, 4, 5, and 6 each had a flow rate of 600 gallons per minute. Pumps 7 and 8 each had a flow rate of 400 gallons per minute.

Pump failures occurred as scripted, according to level of engagement in the condition. In the high engagement conditions, there were 12 failures per 10- minute block (6 per 5-minute block). In the low engagement conditions, there were 6 failures per 10 minute block (3 per 5-minute block). In all cases, pump failures were separated from signals occurring in the system monitoring task by a minimum of 10 seconds. Pump failures were indicated by the illumination of a red light in the malfunctioning pump, which signified that particular pump was unable to transfer fuel. The subject was unable to correct the failure, but was instructed to reorganize his/her strategy until the failure became operational. Once the pump became operational, it was automatically set at "off," and the subject was required to reactivate the pump if he/she wished to use it. If a tank was filled to capacity, the incoming pumps would shut off. Likewise, if a tank was emptied, all outgoing pumps were shut off.

In the automated mode, an "AUTO ON" sign was lighted in the lower right hand corner of the resource (fuel) management window. Scripts operated the resource (fuel) management task, without input from the subject. Typically, the automation would detect and correct a pump failure within 60 seconds. During an instance of "suboptimal" automation, on the other hand, it would take 90 seconds for the pump failure to be corrected. The subject was informed of this possible deviation prior to each session, as well as the fact that he/she may be asked to report such failures in automation performance at a later time. In the case of total automation failure, no pumps would be activated or deactivated, and the "AUTO ON" sign would be extinguished. Each subject was instructed that in the case of total automation failure, he/she was to regain manual control.

Performance measures for the resource (fuel) management task included the following:

(Note: The computer recorded all pump activations, the levels of fuel in Tanks A, B, C, and D at the time of each activation or failure and every 10 seconds.)

. Mean Root Mean Square Errors (RMSE's) were computed for the 10-second intervals (in order that the number of measures would be consistent across conditions and subjects) and averaged across 5-minute blocks for the entire 30-minute session.

The following equation was used to compute RMSE's:

$$\mathbf{RMSE} = \sqrt{[\sum_{i}^{N} \{\Delta\, TA^2 + \Delta\, TB^2\} / N]}$$

where TA and TB are the fuel deviations in tanks A and B, respectively, and N is the sample size (# of readings per 5-minute block).

As in the system monitoring task, this meant that for the conditions under which resource (fuel) management was under manual control, all 6 blocks of measures (for that task) were examined in the analysis. On the other hand, under conditions in which the resource (fuel) management was automated, only blocks 5 and 6 of measures (for that task) would be meaningful to the analysis. Again, the effects of automating the resource (fuel) management task on the remaining two tasks (system monitoring and tracking) could still be examined across all 6 blocks of measures of those tasks.

Tracking

The tracking task was displayed in the upper middle section of the terminal screen. This task was similar to those of the primary flight control of an airplane. A joystick was used to control position on the screen.

The overall purpose of this task was to keep the airplane, represented by a green circle, over the cross-hairs within the dotted rectangular area in the center of the display. If the subject did not control the plane with the joystick, the plane would drift away from the center. Use of the joystick was described in the "Subject Briefing Package" and demonstrated during the training session.

Note: RMSE's were collected by the computer for the tracking task approximately every second. The following excerpt was taken from Parasuraman, Bahri, and Molloy (1991):

Operator performance of the tracking task is evaluated by sampling the x and y control inputs at 10 Hz and thus deriving the x and y deviations. The root mean square (RMS) error is then computed for the samples obtained over a 1-sec period. In computing the combined horizontal and vertical deviations from the target, vertical deviations are converted (in proportion to the monitor x and y resolution) to horizontal pixel units before combination with the horizontal deviations. (p. 10)

Performance measures used in the tracking task included the following:

Mean Root Mean Square Errors (RMSE's) for the 1-second intervals were averaged across 5-minute blocks for the entire 30-minute session. As the Tracking Task was always under manual operation, all 6 blocks of data were at all times meaningful in analysis.

The equation used to compute RMSE was the following:

$$\mathbf{RMSE} = \sqrt{\left[\sum_{i}^{N} \{\Delta xi^2 + (K \ \Delta yi^2)\} \ /N \right]}$$

where x and y are the x and y deviations, K is the monitor resolution ratio (horizontal/vertical), and N is sample size.

Experimental Design

Experiment I

The first investigation included 2 levels of the Between subjects variable Judgment Type (Absolute vs. Comparative) by 3 levels of the within subjects variable Task Type (Automated System Monitoring [ASM] vs. Automated Resource [Fuel] Management [ARM] vs. Fully Manual Control) by 2 levels of the within subjects variable Engagement (High Engagement vs. Low Engagement) in a mixed between/within factorial design (see Table 1).

Note: There were three control conditions, due to the nature of the high/low combinations for the task types. Whenever a particular task was automated, the level of engagement of the "opposing," manual task remained constant. Therefore, there were actually three different combinations of levels of engagement for the task type variable:

1. Automated Resource (Fuel) Management (ARM): High Resource Management (RM)/High System Monitoring (SM) and Low RM/High SM.
2. Automated System Monitoring (ASM): High SM/High RM and Low SM/High RM.

The following controls conditions were provided for each of these:

3. Full Manual Controls: High SM/High RM, High SM/Low RM, and Low SM/High RM.

Table 1: Exp. I: Experimental Design

BETWEEN SUBJECTS	WITHIN SUBJECTS		
JUDGMENT TYPE	GROUP NAME	AUTOMATION/TASK TYPE	ENGAGEMENT LEVEL
Absolute	ARMHH	Res (Fuel) Mgmt	Hi RM, Hi SM
	ARMLH	Res (Fuel) Mgmt	Lo RM, Hi SM
	ASMHH	System Monit	Hi RM, Hi SM
	ASMHL	System Monit	Hi RM, Lo SM
	CONHH	Full Manual	Hi RM, Hi SM
	CONLH	Full Manual	Lo RM, Hi SM
	CONHL	Full Manual	Hi RM, Lo SM
Comparative	ARMHH	Res (Fuel) Mgmt	Hi RM, Hi SM
	ARMLH	Res (Fuel) Mgmt	Lo RM, Hi SM
	ASMHH	System Monit	Hi RM, Hi SM
	ASMHL	System Monit	Hi RM, Lo SM
	CONHH	Full Manual	Hi RM, Hi SM
	CONLH	Full Manual	Lo RM, Hi SM
	CONHL	Full Manual	Hi RM, Lo SM

Experiment II

The second experiment involved the collection of situational awareness data via the SAGAT. The design is displayed in Table 2. It duplicated the first experiment, with the exception that the level of engagement was not examined. Each subject received similar instructions to those in the first experiment, with regards to suboptimal automation and the possibility of automation failure (although subjects in the second experiment never experienced the actual automation failure). Each subject in the second experiment was informed that at some point between 15 and 30 minutes into the simulation, the program would suddenly stop, at which time he/she would be given the questionnaire.

The scripts used to run the conditions included only high engagement levels of both the System Monitoring and the Resource (Fuel) Management Tasks. Six different scripts were generated: two for Automated System Monitoring, two for Automated Resource (Fuel) Management, and two for the fully manual controls. These scripts terminated at intervals between 17 and 27 minutes, and were randomly presented to subjects.

Procedure

Experiment I

The procedure for the first investigation consisted of one training session, followed by the seven conditions: four experimental conditions, and three control (fully manual; one for each engagement combination) conditions. Prior to the investigation, each

Table 2: Exp. II. Experimental Design

BETWEEN SUBJECTS	WITHIN SUBJECTS	
JUDGMENT TYPE	GROUP NAME	AUTOMATION/TASK TYPE
Absolute	ARM	Res (Fuel) Management
	ASM	System Monitoring
	CON	Full Manual
Comparative	ARM	Res (Fuel) Management
	ASM	System Monitoring
	CON	Full Manual

subject was given written instructions (as per Appendix B) concerning how to perform the various tasks. During the first laboratory session, each subject was trained in manual and semi-automated (when either the resource (fuel) management or the system monitoring task was automated) operation of the overall task. This training period began by interviewing the subject to assess his/her knowledge of the "Subject Briefing Package." If the experimenter was satisfied that the subject had read the package, and understood the information to a degree expected on the basis of reading alone (without training), the training session continued. The subject was first shown the animated display of the three tasks on the computer terminal screen (in order to give subjects a visual, animated picture of what they had read). The "hands-on" training then began with the random presentation of one of the three tasks (system monitoring, resource management, or tracking). The manual and automated operation of the task was first described verbally to the subject (as per Appendix F: Task Training Protocol). Upon indication that the subject understood the verbal descriptions, he/she then practiced manual operation of the task for five minutes. This process was repeated for each of the remaining two tasks. The remainder of the training session was divided into three, 5-minute sessions. During the first 5 minute session, the subject practiced manual operation of the tracking task in conjunction with either the system monitoring or the resource (fuel) management task, as determined by random selection. The subsequent 5-minute session then entailed the practice of manual operation of the tracking task in conjunction with the task which had not been practiced in the preceding session. This was done to simulate conditions in which either the system monitoring or the resource (fuel) management task would be automated, and the subject would be manually performing the tracking task in conjunction with the remaining manual task. The final 5-minute session required the practice of all three tasks simultaneously (to simulate fully manual control conditions). Each subject was required to master the practice session by meeting the "passing requirements" described in Appendix F.

Upon completion of the training session, the subject was given a 5-minute break. Following the break, the subject was presented with three successive conditions (experimental and/or control, as determined by random selection). Each of the conditions lasted 30 minutes.

In the case of experimental conditions (in which either System Monitoring or Resource (Fuel) Management was automated), an automation failure occurred 20 minutes into the session. During the remaining 10 minutes, all tasks were then to be performed manually.

Immediately following completion of an experimental session, the subject was queried about the functioning of the automation (except in the control conditions, where no automation was involved). Immediately following the query, the subject was presented the NASA-TLX.

Each subject was instructed to return at the same time the following day. The second day began with a 15-minute practice session: 5 minutes manual operation of Tracking and System Monitoring, 5 minutes manual operation of Tracking and Resource (Fuel) Management (the order of which was determined randomly) and 5 minutes manual operation of all three tasks simultaneously. Following a 5-minute break, the remainder of the subject run proceeded as per the first day, with the exception the subject underwent four, as opposed to three successive conditions (separated by a 5-minute break between the 2nd and 3rd conditions).

Experiment II

The procedure for the second investigation consisted of one training session, four experimental sessions (two sessions in which System Monitoring was automated, and two sessions in which Resource [Fuel] Management was automated), and two control sessions. There were two sessions for each of the two experimental sessions and the control session in order to increase the measures of situational awareness. As indicated in the design, Level of Engagement was not manipulated. Both the System Monitoring and the Resource (Fuel) Management Task always operated at high engagement levels.

As in the first experiment, each subject was given, prior to the investigation, written instructions (Appendix C) concerning how to perform the various tasks. The training session for the second experiment was identical to that in the first experiment, with respect to operation of the tasks. It differed in that subjects were told that at some time between 15 and 30 minutes during a session, the simulation would stop, at which time he/she was to complete the SAGAT questionnaire.

Each subject in the second experiment was told in the training session, and reminded before participating in each condition, of the manner in which he/she was to answer the questionnaire. The subject was told that when the simulation stopped, he/she was to immediately begin answering the questionnaire (located face-down on the table to the right of the computer terminal). The subject was instructed not to view the questionnaire prior to that point in time. Each subject was instructed to:

> Answer all questions as quickly, but as accurately as possible. There is no time limit, but do not spend too much time on any one question... your first response is usually the most accurate. You may return to questions at a later time. If you feel you do not know the answer to a particular question, it is still important that you guess.

Each subject in the second experiment underwent three sessions the first day, and three sessions the second day. On each day, the subject was given a 5-minute break between the 2nd and 3rd sessions. The order of these sessions was randomized.

CHAPTER III

RESULTS

Study I Primary Analysis: Performance Data

For each subject, a mean was computed per dependent variable per task per 5-minute block (in the case of reaction time in the System Monitoring Task, the median was computed). These block means were entered into Complete Statistical System (CSS) 3.1 statistical package (Statsoft, 1992). All percent correct data were first subjected to an arcsine transformation. Several separate ANOVAs were conducted. Analyses focused upon Main Effects and Interactions of Judgment Types (Absolute vs. Comparative) and "Conditions." "Condition" comprised the Task Type-Engagement Level combination. Therefore, there were seven Conditions:

(1) Automated Resource (Fuel) Management; High Engagement Level Resource (Fuel) Management / High Engagement Level System Monitoring (**ARMHH**).

(2) Automated Resource (Fuel) Management; Low Engagement Level Resource (Fuel) Management / High Engagement Level System Monitoring (**ARMLH**).

(3) Automated System Monitoring; High Engagement Level Resource (Fuel) Management / High Engagement Level System Monitoring (**ASMHH**).

(4) Automated System Monitoring; High Engagement Level Resource (Fuel) Management / Low Engagement Level System Monitoring (**ASMHL**).

(5) Full Manual CONtrol; High Engagement Level Resource (Fuel) Management / High Engagement Level System Monitoring (**CONHH**).

(6) Full Manual CONtrol; Low Engagement Level Resource (Fuel) Management / High Engagement Level System Monitoring (CONLH).

(7) Full Manual CONtrol; High Engagement Level Resource (Fuel) Management / Low Engagement Level System Monitoring (CONHL).

The Greenhouse-Geisser (1958) correction was applied to all significant effects involving repeated measures with more than two levels. This procedure was carried out to compensate for possible violations in the assumptions of compound symmetry (homogeneity of variances within-group and covariances across subjects) or sphericity (that the model for within-subjects consists of orthogonal elements).

Tracking Task: Blocks 1-4

During the first twenty minutes (Blocks 1-4) of each trial, either the system monitoring or the resource (fuel) management task was automated, except in the case of the full manual controls, where all three tasks were manually performed by the subjects for the entire session. A 3-Way Analysis of Variance (ANOVA) was performed on the dependent variable for tracking, root mean square error (RMSE). The three factors entered into the analysis were judgment type (JT) (absolute vs. comparative), block number (Blk) (1, 2, 3, 4) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine the possible effects of automating either the system monitoring task or the resource (fuel) management task on tracking performance.

Table 3 displays the summary analysis for the tracking RMSE during blocks 1-4 of the tracking task across judgment types, block numbers, and conditions. As indicated

Table 3: Summary Analysis for Tracking RMSE Blocks 1-4

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3093.77	14	13165.33	0.235	0.635
Blk	3	10053.18	42	902.46	11.140	0.000
Cond	6	494.61	84	389.13	1.271	0.279
JTxBlk	3	1836.51	42	902.46	2.035	0.124
JTxCond	6	708.10	84	389.13	1.820	0.105
BlkxCond	18	1444.30	252	391.91	3.685	0.000
JTxBlkx Cond	18	774.93	252	391.91	1.977	0.011

in the summary table, there was a significant main effect for block number, $\underline{F}(3, 42) = 11.14$, $p < .01$. There was also a significant two-way interaction between block number and condition, $\underline{F}(18, 252) = 3.68$, $p < .01$. Finally, there was a significant three-way interaction between judgment type, block number, and condition. However, when the Greenhouse-Geisser procedure was applied to the three-way interaction, the procedure adjusted the degrees of freedom for the effect from 18 to 2.68 and it adjusted the degrees of freedom for error from 252 to 37.46. This resulted in a loss of significance, adjusting the associated p-level from p = .012 to p = .134.

The two-way interaction between block number and condition, on the other hand, remained significant following the Greenhouse-Geisser procedure. The Greenhouse-Geisser adjusted the degrees of freedom for the effect from 18 to 2.676 and corrected the degrees of freedom for the error from 252 to 37.46. The p-value for the two-way interaction corrected from p = .000 to p = .020.

Considering the main effect for block number, it can be seen in Figure 4, which displays the means and standard errors for the block numbers, that tracking performance was best in the first five minutes of the session (block 1). An LSD post hoc test was applied to the main effect for block number. This test revealed that Tracking RMSE in block 1 was significantly lower than in blocks 2 (p = .000005), 3 (p = 00006) and 4 (p = .0001). Blocks 2, 3 and 4 were not significantly different from each other.

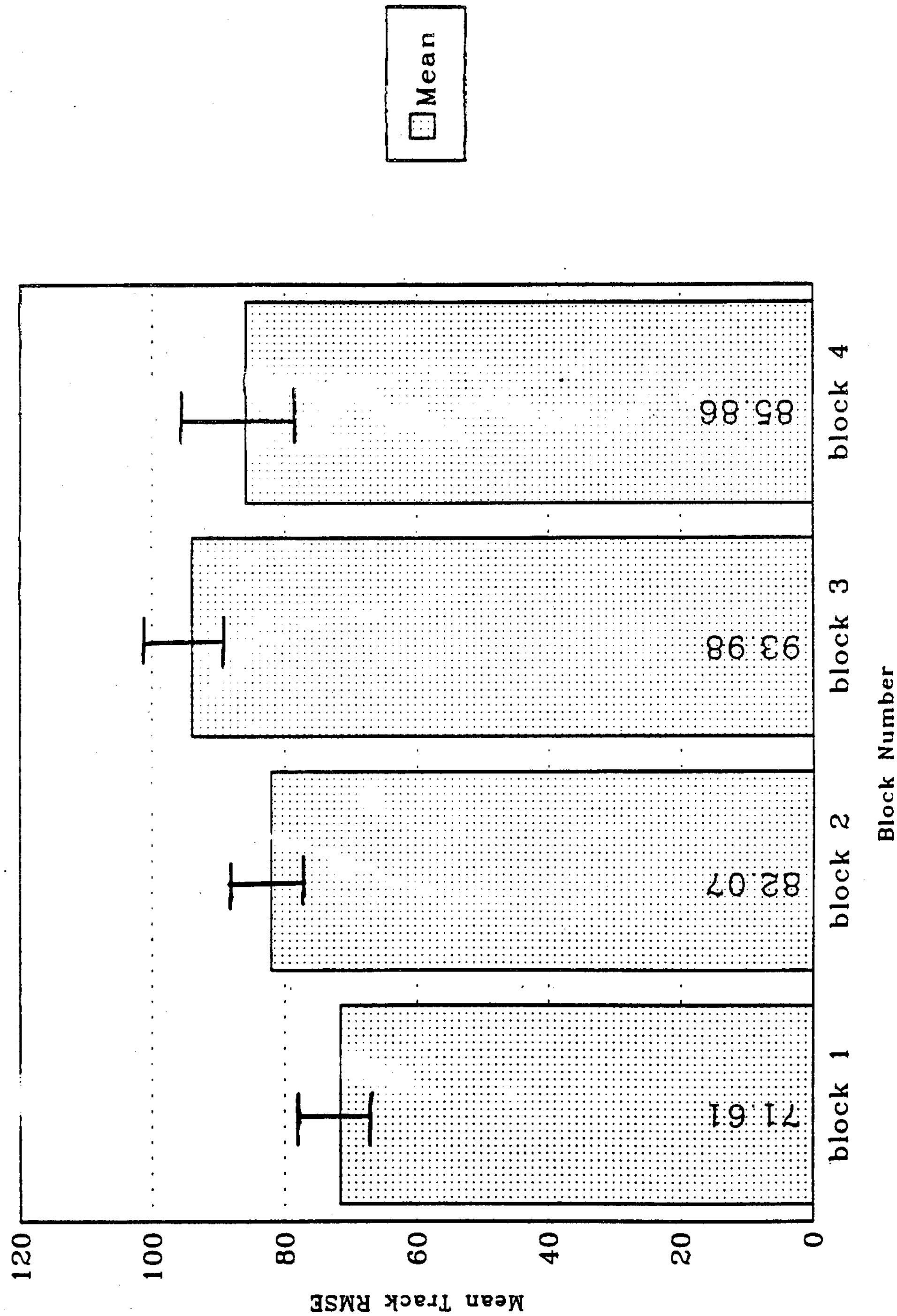


Figure 4: Block Means
Tracking RMSE Blocks 1-4

In order to interpret the two-way interaction between block number and condition, four separate analyses were conducted. These included tracking RMSE for block 1 across all seven conditions, tracking RMSE for block 2 across all seven conditions, tracking RMSE for block 3 across all seven conditions, and tracking RMSE for block 4 across all seven conditions. Significant effects were found only in the analyses for blocks 2 and 3.

Table 4 displays the summary analysis for tracking RMSE in block 2 of the tracking task across all seven conditions (ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). As indicated in the summary table, there was a significant main effect for condition, F (2, 84) = 7.28, $p < .01$. There was also a significant two-way interaction between judgment type and condition, F (6, 84) = 2.66, $p < .05$. A Greenhouse-Geisser correction was applied to both the main effect for condition and the two-way interaction between judgment type and condition. In the case of the main effect for condition, the Greenhouse-Geisser procedure adjusted the degrees of freedom for the effect from 6 to 3.52, and it adjusted the degrees of freedom for error from 84 to 49.24. The main effect remained significant, with the p-level adjusting from p = .020 to p = .044.

In the case of the two-way interaction between judgment type and condition, the Greenhouse-Geisser corrected the degrees of freedom for the effect from 6 to 3.52 and it corrected the degrees of freedom for error from 84 to 49.24. The two-way interaction also remained significant, correcting from p = .020 to p = .044.

Table 4: Summary Analysis for Tracking RMSE Block 2

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	12.70	14	2970.89	0.075	0.789
Cond	6	3981.95	84	267.09	7.281	0.000
JTxCond	6	735.44	84	267.09	2.658	0.021

In examining the means and standard errors for the main effect for condition, Figure 5 shows that Tracking performance was better (lower RMSE) while the Resource (Fuel) Management Task was automated. An LSD post-hoc analysis revealed that Tracking performance was significantly better under the condition in which Fuel Management was automated (ARMHH), as compared to when System Monitoring was automated (ASMHH), $p = .000001$ and as compared to its full manual control condition (CONHH), $p = .0001$. Tracking performance was also significantly better under the condition ARMLH than under its full manual control condition (CONLH), $p = .048$.

In examining the means and standard errors for the two-way interaction between judgment type and condition, Figure 6 shows that subject performance on the Tracking Task was generally better under the Comparative, as opposed to Absolute judgment types, except when System Monitoring was automated. An LSD post-hoc analysis again revealed that Tracking performance was better (lower RMSE) when Resource (Fuel) Management was automated. In the case of absolute judgment types, tracking performance under the condition ARMHH was significantly better than under ASMHH ($p = .042$) or under CONHH ($p = .015$). Additionally, ARMLH was significantly better than CONLH ($p = .022$). Similarly, in the case of comparative judgment types, tracking performance under ARMHH was significantly better than under ASMHH ($p = .000001$) or under CONHH ($p = .001$). However, there were no significant differences in tracking performance between absolute and comparative judgment types for automated resource (fuel) management. In other words, absolute ARMHH did not

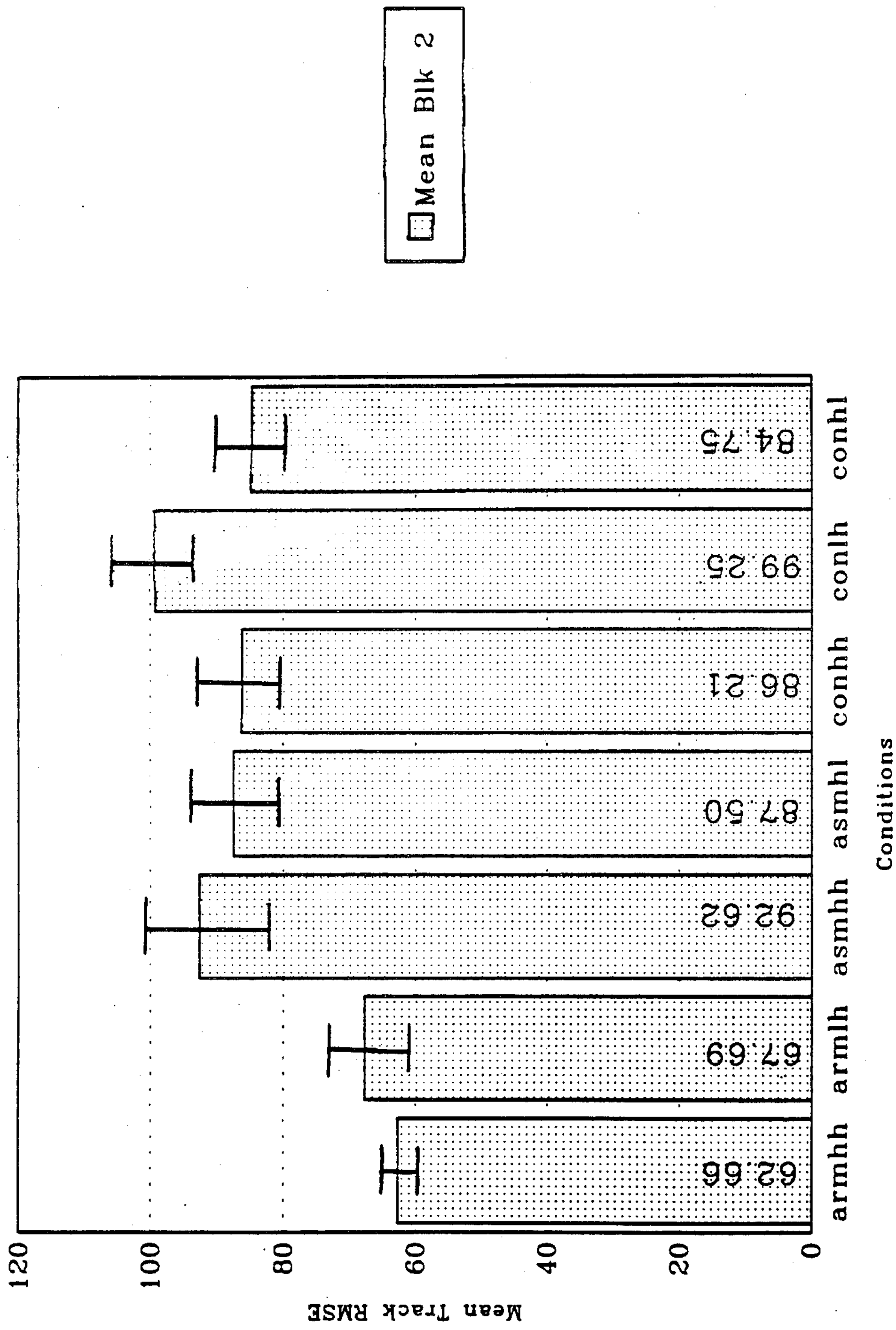


Figure 5: Condition Means
Tracking RMSE Block 2

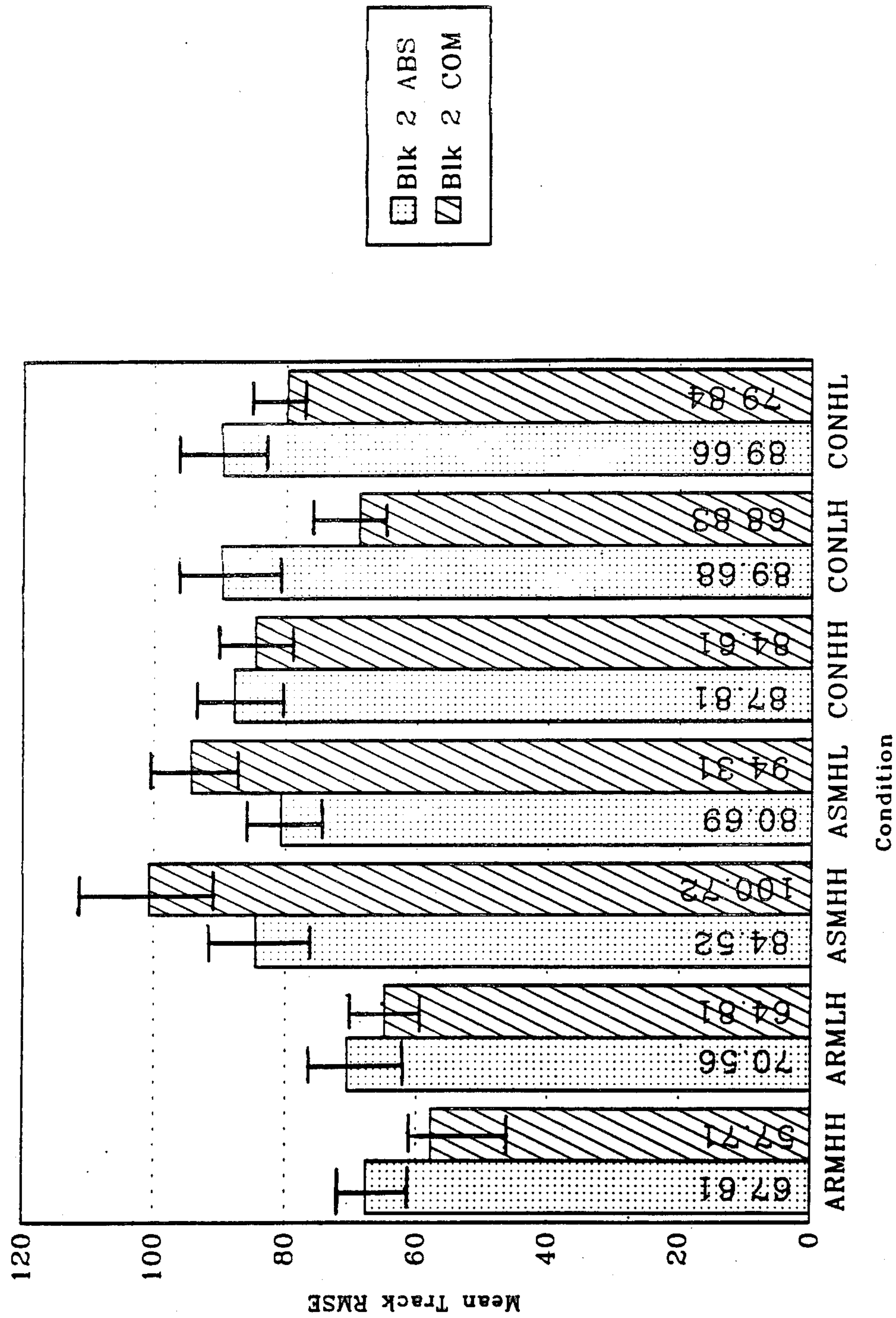


Figure 6: Judgment Type x Condition Means Tracking RMSE Block 2

differ significantly from comparative ARMHH, nor did absolute ARMLH differ significantly from comparative ARMLH. However, under the comparative judgment type, a possible effect due to the engagement level of the resource (fuel) management task did emerge. Tracking performance under the control condition in which the engagement level of the resource (fuel) management task and the engagement level of the system monitoring task were both high (CONHH) was worse than the control condition in which the engagement level of the resource (fuel) management task was low, while the engagement level of the system monitoring task remained high (CONLH). This difference approached significance at $p = .057$.

While there were no significant differences between absolute and comparative judgment types under conditions where resource (fuel) management was automated, tracking performance was consistently better (lower RMSE) under the comparative judgment type, as would be expected. Likewise, in the case of the full manual control conditions, tracking performance under comparative CONLH was significantly better than under absolute CONLH ($p = .012$). The opposite was seen, however, in the case of absolute versus comparative judgment types in conditions where system monitoring was automated. When system monitoring was automated, tracking performance was actually better under absolute judgment conditions. In fact, tracking performance under absolute ASMHH was significantly <u>better</u> than under comparative ASMHH ($p =$ 050). In the case of the comparative judgment type, tracking performance was actually significantly worse while system monitoring was automated, with subjects performing significantly better under the control condition CONHH as compared to ASMHH ($p = .052$).

In summary, in the case of tracking performance during block 2, subjects performed generally better under comparative judgment types, except when system monitoring was automated. Tracking performance was improved when resource (fuel) management was automated, both under absolute and comparative judgment types, and in the case of the control conditions, tracking performance was possibly aided by a lower engagement level within the resource (fuel) management task. In the case of absolute judgment types, automating system monitoring did not appear to improve tracking performance. However, in the case of comparative judgments types, automating system monitoring actually had a detrimental effect on tracking performance.

The general trends in these findings support Experiment I, Hypothesis #1(a) (p. 27), which states that automation of the resource (fuel) management task was expected to have a greater positive effect on performance of the non-automated tasks. Although there were no specific predictions concerning the effects of judgment type on performance in the tracking task, the fact that tracking performance was consistently (though not significantly) better under the comparative judgment type when resource (fuel) management was automated is in line with Experiment I, Hypothesis # 3 (p. 32). The findings with respect to the effects of judgment type on tracking performance while system monitoring was automated are interesting. Though no specific effects of judgment type on tracking performance were predicted, the most likely explanation for these results is related to Hypothesis # 3b (p. 32), which states that absolute judgment types were expected to impose a higher overall workload across conditions. Perhaps the comparative judgment type reduced workload to a level which, coupled with automation of the system monitoring task, reduced overall task performance.

Table 5 displays the summary analysis for tracking RMSE in block 3 of the tracking task across all seven conditions. As indicated in the summary table, there was a significant main effect for condition, $\underline{F}$ (2, 84) = 6.23, $p < .01$. There was also a significant two-way interaction between judgment type and condition. A Greenhouse-Geisser correction was applied to both the main effect and the interaction. In the case of the main effect for condition, the Greenhouse-Geisser corrected the degrees of freedom for the effect from 6 to 3.57 and it corrected the degrees of freedom for error from 84 to 49.93. The main effect remained significant, correcting from $p = .00002$ to $p = .00038$. In the case of the two-way interaction between judgment type and condition, the Greenhouse-Geisser corrected the degrees of freedom for the effect from 6.0 to 3.57 and it corrected the degrees of freedom for error from 84 to 49.93. The two-way interaction also remained significant, correcting from $p = .022$ to $p = .046$.

It can be seen in Figure 7, which displays the means and standard errors for the main effect for condition, that tracking performance in block 3 was better (lower RMSE) while the Resource (Fuel) Management Task was automated. An LSD post-hoc analysis revealed that tracking performance was significantly better under the condition ARMHH, as compared to when System Monitoring was automated

Table 5: Summary Analysis for Tracking RMSE Block 3

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1163.64	14	4371.81	0.266	0.614
Cond	6	1981.65	84	318.08	6.23	0.000
JTxCond	6	831.75	84	318.08	2.61	0.023

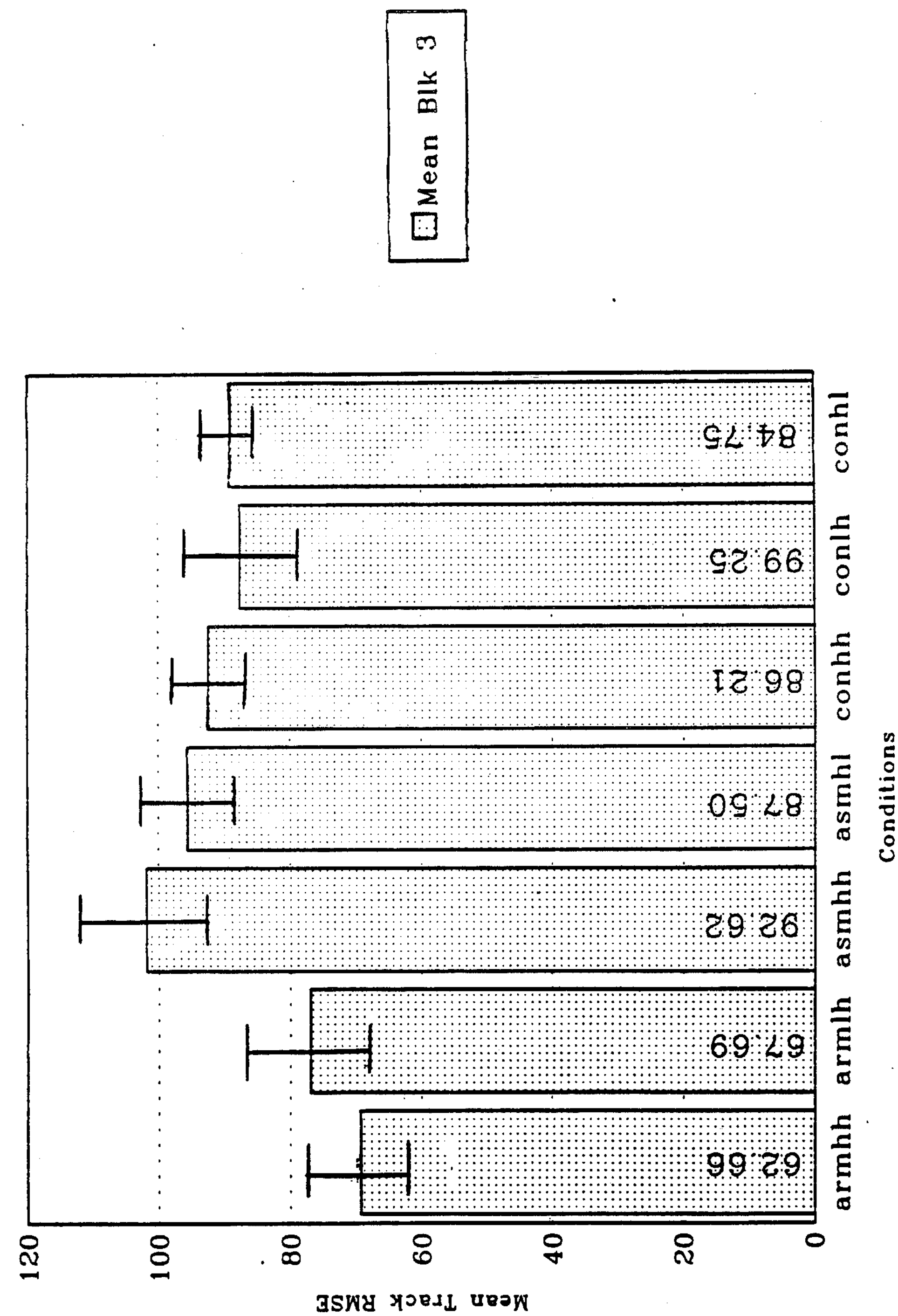


Figure 7: Condition Means
Tracking RMSE Block 3

(ASMHH), p = .000002 and as compared to its full manual control condition (CONHH), p = .0004.

Two distinct findings are highlighted in Figure 8, which displays the means and standard errors for the two-way interaction between judgment type and condition. First, tracking performance in block 3 was generally better (lower RMSE) under the comparative judgment types. Second, tracking performance in block 3 was generally better while Resource (Fuel) Management was automated.

With respect to the first general finding, that tracking performance was better under comparative conditions, an LSD post-hoc test revealed that tracking performance was significantly better under the condition ARMHH in the comparative judgment type than under the same condition in the absolute judgment type (p = .02). Furthermore, tracking performance was significantly better under the control condition CONLH in the comparative judgment type than under the same condition in the absolute judgment type (p = .01).

With respect to the second general finding, that tracking performance was better while resource (fuel) management was automated, an LSD post-hoc test revealed that, under the comparative judgment type, tracking performance was significantly better under ARMHH than under ASMHH (p = .000), or CONHH (p = .001).

In summary, the greatest benefits in tracking performance during block 3 occurred when the resource (fuel) management task was automated and the judgment type was

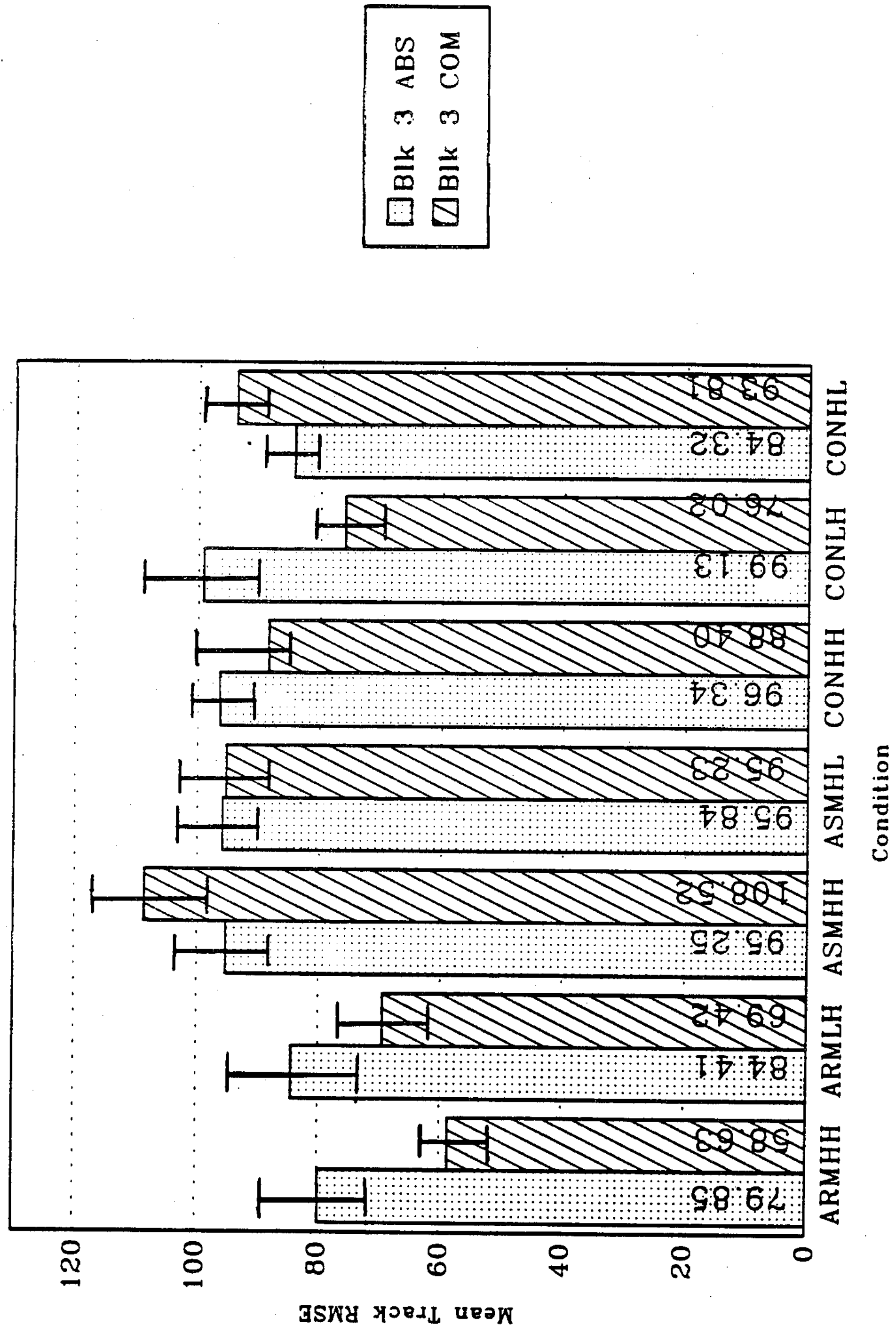


Figure 8: Judgment Type x Condition Means Tracking RMSE Block 3

comparative. As in block 2, the finding that automation of the resource (fuel) management task aided tracking performance supports Experiment I, Hypothesis # 1(a) (p. 27).

Tracking Task: Block 5

Block 5 of each session encompassed the first 5 minutes following automation failure of either the system monitoring or the resource (fuel) management task (except in the full manual control conditions, in which all three tasks were manually performed by the subject the entire session). Recall that subjects were instructed that upon detecting an automation failure, they were to regain manual control of the previously automated task, as well as to continue manual control of the other two tasks.

A 2-Way ANOVA was performed on tracking RMSE for block 5. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine the possible effects of automation failure in the system monitoring or the resource (fuel) management task on subsequent performance on the tracking task.

Table 6 displays the summary analysis for tracking RMSE in block 5 across judgment types and conditions. As indicated in the summary table, there were no significant main effects or interactions for block 5 tracking performance. Tracking performance did not change significantly across judgment types or conditions in block 5. No predictions were made with respect to the effects of automation failure (which

Table 6: Summary Analysis for Tracking RMSE Block 5:

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3444.11	14	3514.76	0.98	0.34
Cond	6	451.57	84	445.75	1.01	0.42
JTxCond	6	565.34	84	445.75	1.27	0.28

precedes block 5, except in the full manual control conditions) on tracking performance. The finding indicates that the failure of automation in either the system monitoring or resource (fuel) management task had no effect on tracking performance. An examination of the means for the effects reveals no particular pattern. The means, standard deviations, and standard errors of all effects are displayed in Appendix G.

Tracking RMSE: Block 6

Block 6 encompassed the final five minutes of each session. A 2-Way ANOVA was performed on tracking RMSE for block 6. The two factors entered into the analysis were Judgment Type (JT) (Absolute vs. Comparative) and Condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine any lingering effects of automation failure of the system monitoring or the resource (fuel) management task on subsequent performance on the tracking task.

Table 7 displays the summary analyses for Tracking RMSE in block 6. As indicated in the summary table, there were no significant main effects or interactions for block 6 Tracking RMSE. Tracking performance did not change significantly across judgment types or conditions in block 6.

Tracking Task: Blocks 1-6

In order to examine any general trends in tracking performance over an entire condition session, a separate 2-way ANOVA was performed on tracking RMSE for blocks 1-6 for each of the seven conditions. The two factors entered into the analysis

Table 7: Summary Analysis for Tracking RMSE Block 6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3247.91	14	4948.86	0.66	0.43
Cond	6	217.31	84	404.25	0.54	0.78
JTxCond	6	259.65	84	404.25	0.64	0.70

were judgment type (JT) (absolute vs. comparative) and block number (Blk) (1, 2, 3, 4, 5, 6). Of the seven conditions examined, all had significant main effects for block (and one for a two-way interaction between judgment type and block). However, following Greenhouse-Geisser corrections on all significant findings, two of the conditions, ARMLH and ASMHH lost significance. There was a general trend across all seven conditions, in which tracking performance declined, overall, through block 6. The specific results of the separate analyses follow.

Considering the condition ARMHH, Table 8 displays the summary analysis for tracking in blocks 1-6. As indicated, there was a main effect for block number, $\underline{F}$ (5, 70) = 17.10, $p < .01$. Following Greenhouse-Geisser, the degrees of freedom for effect adjusted from 5 to 1.89 and the degrees of freedom for error adjusted from 70 to 26.51. The p-value remained significant, adjusting from $p = .00000$ to $p = .00002$.

Figure 9 shows a consistent decline in tracking over the 6 blocks. Post-hoc LSD analyses indicated block 1 performance was significantly better than blocks 3 ($p = 0.045$), 4 ($p = 0.001$), 5 ($p = 0.000$), and 6 ($p = 0.000$). Likewise, performance in block 2 was significantly better than blocks 4 ($p = 0.003$), 5 ($p = 0.000$), and 6 ($p = 0.000$). Performance in blocks 4 and 5 did not differ from each other significantly, but both differed significantly from block 6, ($p = 0.000$) and ($p = .000$), respectively. Considering ARMLH, Table 9 displays the summary analysis for tracking blocks 1-6. There was a main effect for block number, $\underline{F}$ (5, 70) = 3.06, $p < .05$. With the Greenhouse-Geisser procedure, the degrees of freedom for effect adjusted from 5 to

Table 8: Summary Analysis for ARMHH Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	4344.48	14	2549.09	1.704	0.213
Cond	5	2021.24	70	118.18	17.103	0.000
JTxCond	5	216.44	70	118.18	1.831	0.118

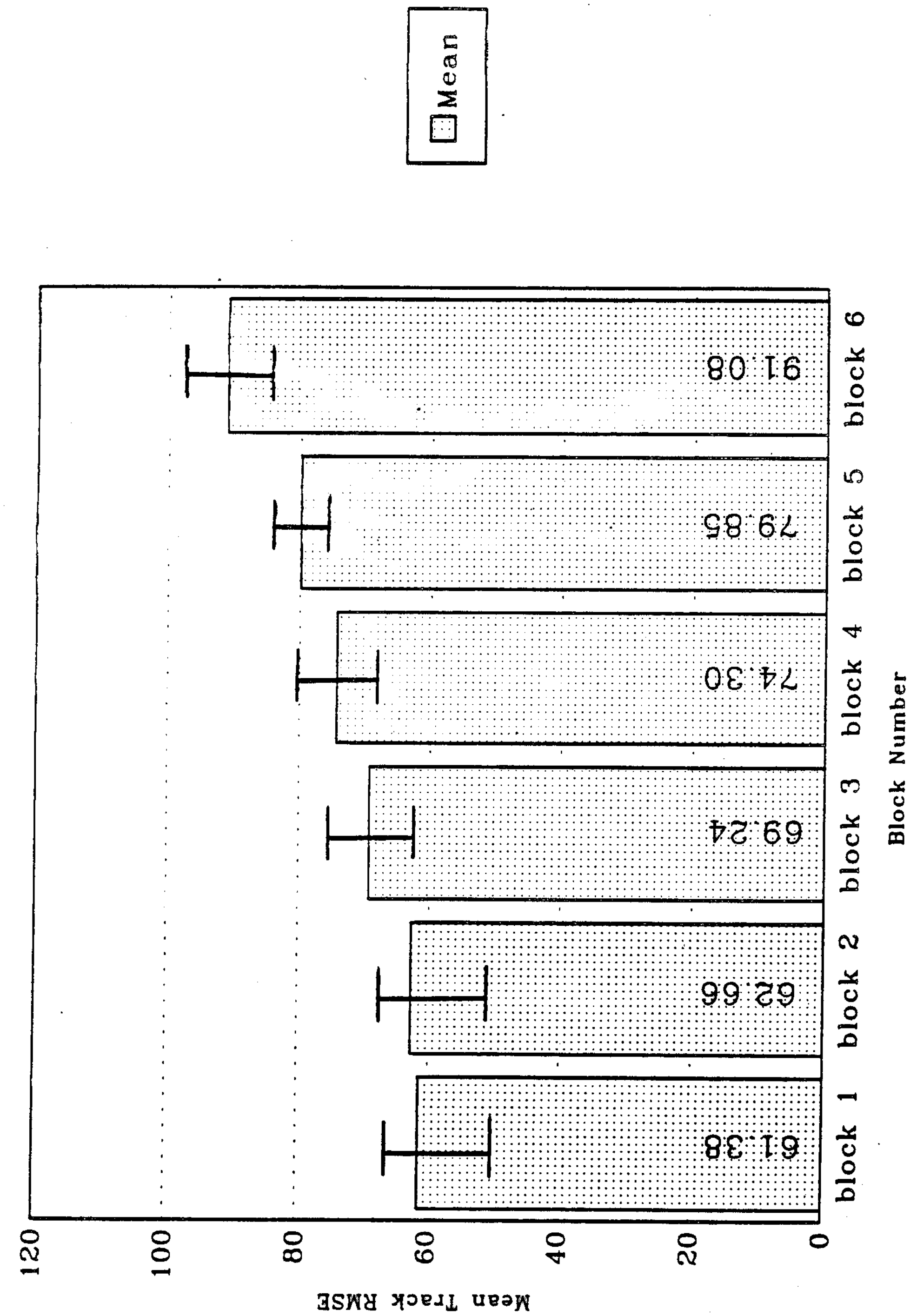


Figure 9: ARMHH Block Means
Tracking RMSE Blocks 1-6

Table 9: Summary Analysis For ARMLH Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	7986.75	14	4664.23	1.712	0.212
Cond	5	2071.99	70	677.42	3.059	0.015
JTxCond	5	727.68	70	677.42	1.074	0.382

1.61 and the degrees of freedom for error were corrected from 70 to 22.59. The p-value lost significance, with a correction from p = .015 to p = .066.

Despite the loss of the significant main effect for block, Figure 10, which displays the means and standard errors for the main effect, shows there is a consistent decline in tracking performance over the 6 blocks. This decline is quite similar to that seen in the previously examined condition, ARMHH.

Considering the condition ASMHH, Table 10 displays the summary analysis for tracking RMSE in blocks 1-6. As indicated in the summary table, there was an initial main effect for block number, $\underline{F}$ (5, 70) = 2.38, p < .05. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 1.76 and the degrees of freedom for error were corrected from 70 to 24.64. The p-value lost significance, with a correction from p = .047 to p = .113.

Despite the loss of the significant main effect for block, Figure 11, which displays the means and standard errors for the main effect, shows there is a consistent decline in tracking performance over the 6 blocks. This decline is quite similar to that seen in the previously examined condition, ARMHH.

Considering the condition ASMHL, Table 11 displays the summary analysis for tracking RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, $\underline{F}$ (5, 70) = 5.84, p < .01. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 2.08 and the degrees of freedom for error were corrected from 70 to 29.10. The p-value remained significant, with a correction from p = .00014 to p = .0074. There was also a significant

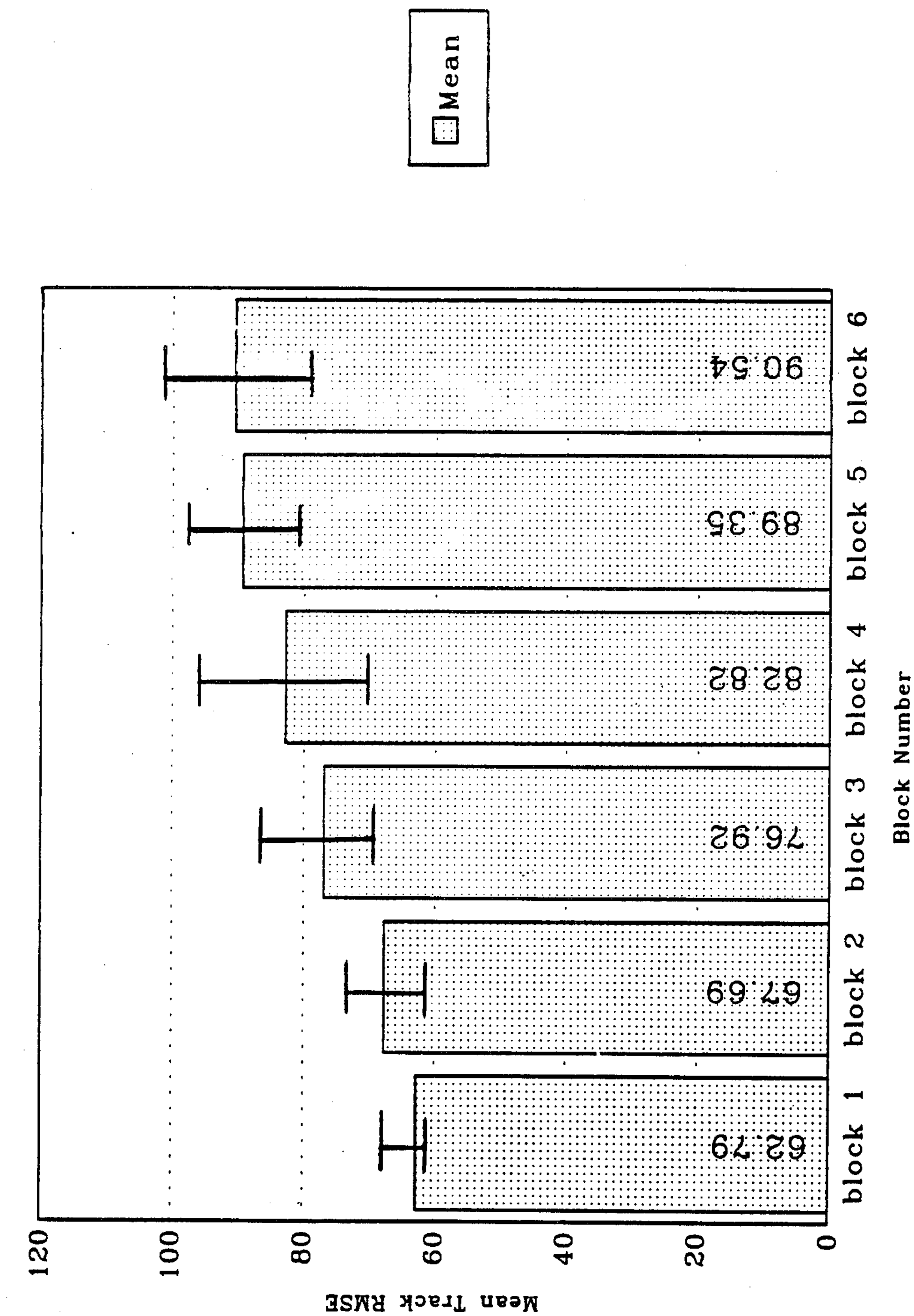


Figure 10: ARMLH Block Means
Tracking RMSE Blocks 1-6

Table 10: Summary Analysis for ASMHH Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1042.14	14	6711.72	0.155	0.699
Cond	5	1958.13	70	820.82	2.386	0.047
JTxCond	5	865.62	70	820.82	1.055	0.393

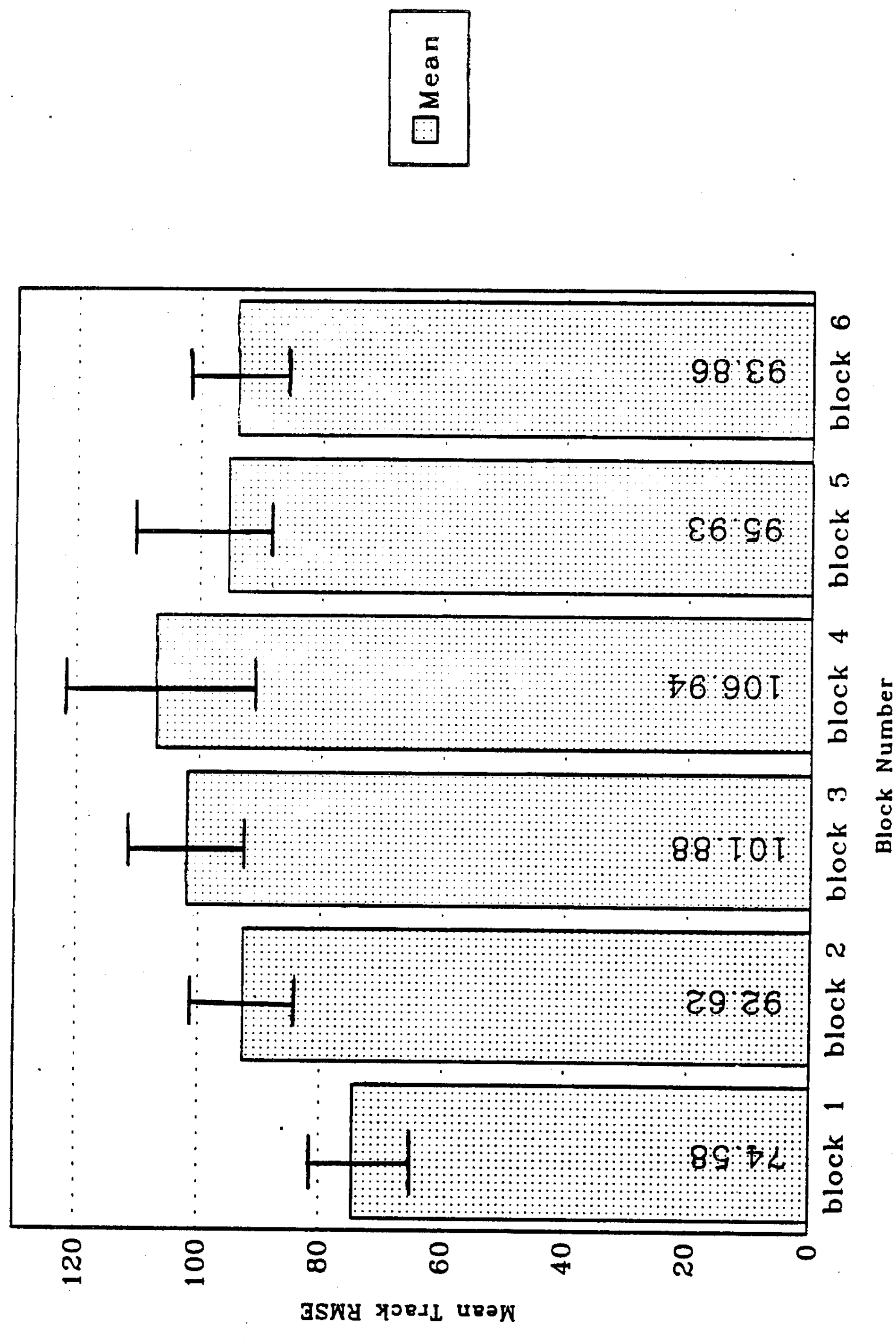


Figure 11: ASMHH Block Means
Tracking RMSE Blocks 1-6

Table 11: Summary Analysis for ASMHL Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3.71	14	3270.94	0.001	0.973
Cond	5	713.32	70	122.13	5.841	0.000
JTxCond	5	654.25	70	122.13	5.357	0.000

two-way interaction between judgment type and block number under ASMHL. It, too, remained significant following the Greenhouse-Geisser procedure, with degrees of freedom for effect corrected from 5 to 2.08, degrees of freedom for error corrected from 70 to 29.10, and the p-value corrected from $p = .00032$ to $p = .010$.

As can be seen in Figure 12, which displays the means and standard errors for the main effect for block, there is a general decline in tracking performance over the 6 blocks, albeit a bit less consistent than in previously discussed conditions. Post-hoc LSD analyses indicated that performance in block 1 was significantly better (lower RMSE) than blocks 2 ($p = .005$), 3 ($p = .000$), 4 ($p = .000$), and 6 ($p = .001$). Likewise, performance in block 2 was significantly better than block 3 ($p = .043$). However, the typical pattern is then broken, with performance in block 3 significantly worse (higher RMSE) than in block 5 ($p = .003$). Performance in blocks 4, 5 and 6 did not differ from each other significantly.

This change in the typical pattern of initial block performance being consistently better than later block performance is made more intriguing upon examination of Figure 13, which displays the means and standard errors for the two-way interaction between judgment type and block number. As shown in the figure, the absolute judgment type conditions demonstrate a fairly consistent drop in tracking performance through block 6, particularly in blocks 5 and 6, which follow the automation failure in the System Monitoring Task. However, in the case of the comparative judgment type conditions, the figure shows that, while there is a consistent drop in performance during blocks 1-4 (i.e., while System Monitoring is automated), there is a drastic

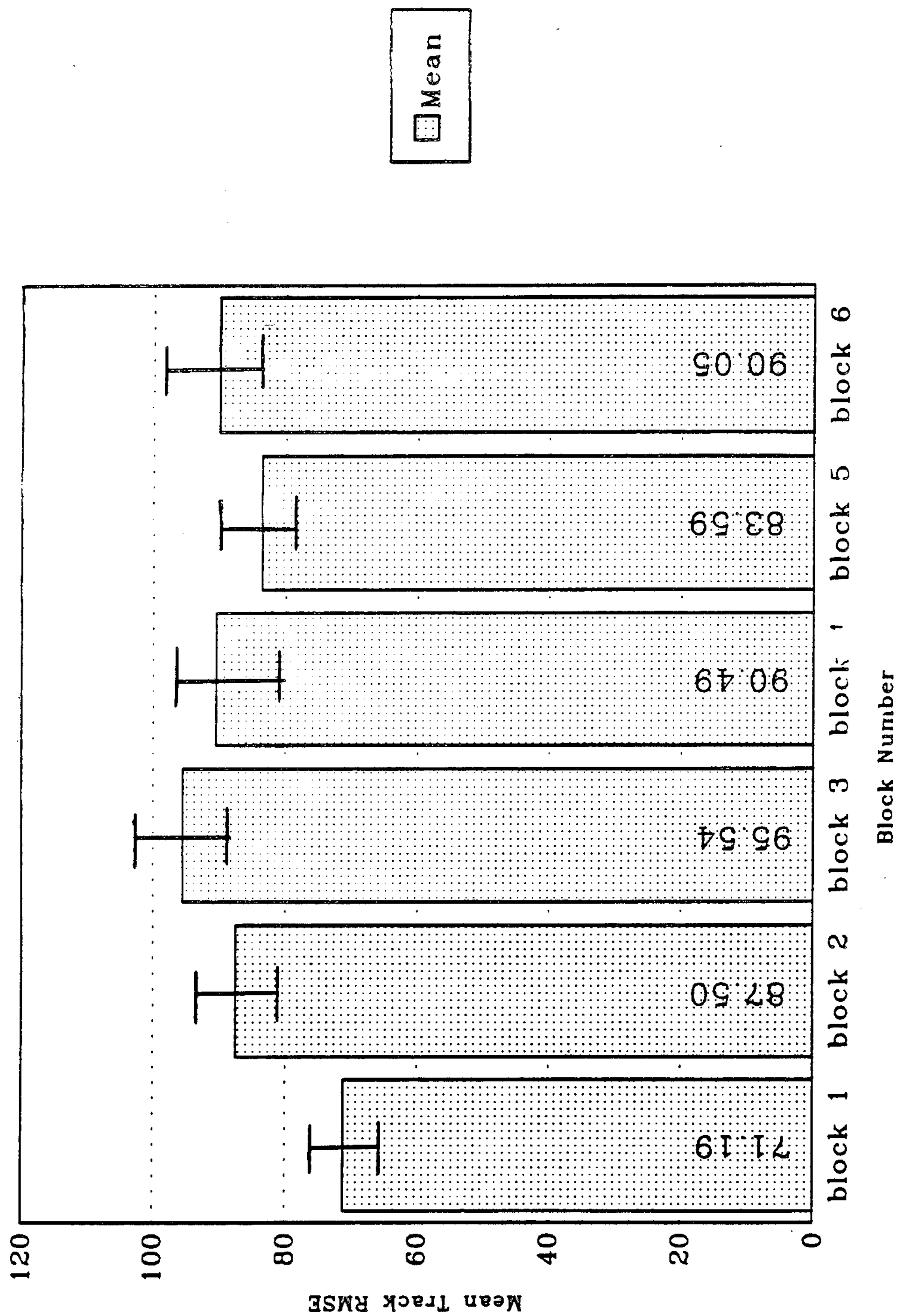


Figure 12. ASMHL Block Means
Tracking RMSE Blocks 1-6

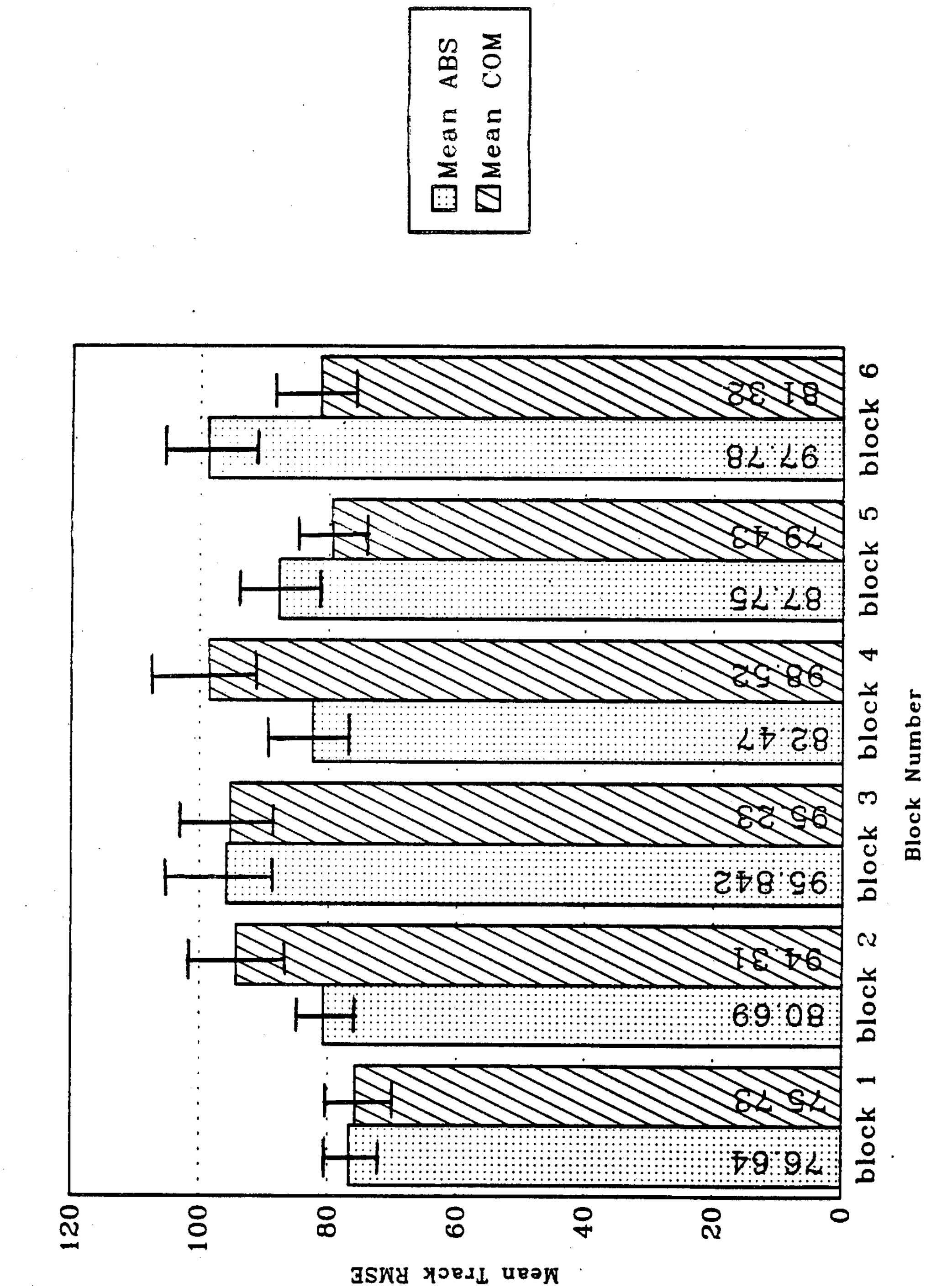


Figure 13: ASMHL Judgment Type x Block Means Tracking RMSE

improvement in performance in blocks 5 and 6, following the automation failure. In relation to Experiment I, Hypothesis # 4 (p. 32), these findings indicate that in both the absolute and comparative judgment types, the characteristic vigilance drops are seen in blocks 1-4. However, performance differences in blocks 5 and 6 may perhaps be attributed to differences between the absolute and comparative judgment types in the stability of the internal model. In the case of the less stable model (absolute judgment type), the confusion resulting from regaining manual control of a more dynamic model task may account for the performance drops in blocks 5 and 6. With the more stable model (comparative judgment type), on the other hand, it appears that detection of automation failure may have served as an overall task "wake-up call," resulting in improved performance in blocks 5 and 6.

Considering the condition CONHH, Table 12 displays the summary analysis for tracking RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, $\underline{F}$ (5, 70) = 10.24, p < .01. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 3.55 and the degrees of freedom for error were corrected from 70 to 49.74. The p-value remained significant, as well as remaining at p = .00000.

Figure 14, which displays the means and standard errors for the main effect for block, shows the most notable features of tracking performance are in blocks 1 and 3. Subjects, as is consistent with previously discussed conditions, performed best in the first block. Post-Hoc LSD analyses indicated that tracking performance in block 1 was, in fact, significantly better (lower RMSE) than in blocks 2 (p = .002), 3 (p =

Table 12: Summary Analysis for CONHH Tracking Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1013.81	14	3303.78	0.307	0.588
Cond	5	732.68	70	71.52	10.244	0.000
JTxCond	5	22.63	70	71.52	0.316	0.902

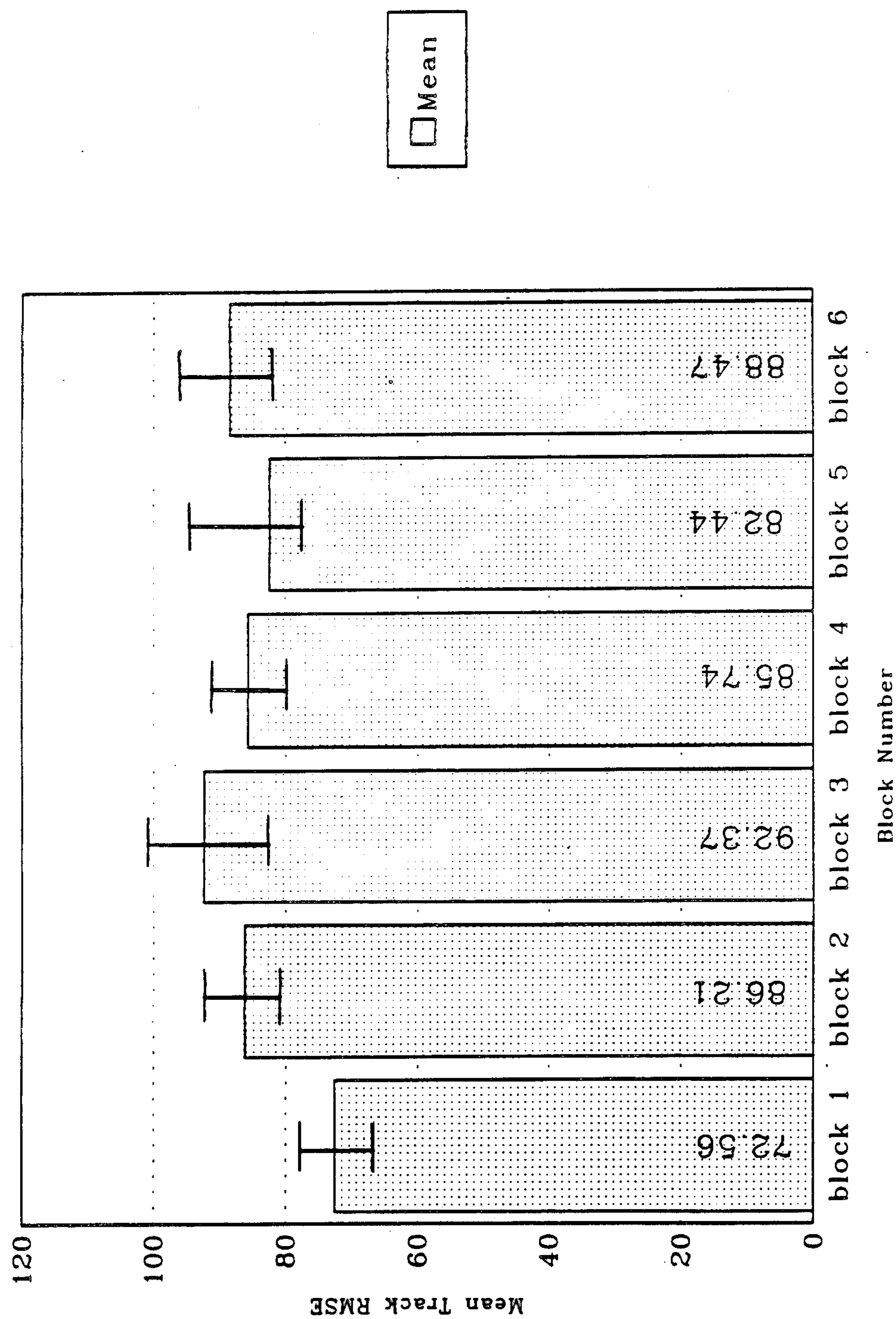


Figure 14 CONHH Block Means
Tracking RMSE Blocks 1 6

.000), 4 (p = .003), 5 (p = .023), and 6 (p = .000). Performance then began to decline in block 2, and reached its worst level of the session in block 3. Post-Hoc LSD analyses indicated that tracking performance in block 3 was significantly worse (higher RMSE) than in blocks 2 (p = .000), 4 (p = .000), 5 (p = .000) and 6 (p = .000). Performance then began to improve in blocks 4 - 5, declining slightly again in block 6, with no significant differences in tracking performance between these blocks.

Considering the condition CONLH, Table 13 displays the summary analysis for tracking RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, $\underline{F}$ (5, 70) = 4.54, $p < .01$. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 2.13 and the degrees of freedom for error were corrected from 70 to 29.80. The p-value remained significant, correcting from p = .002 to p = .019.

Figure 15, which displays the means and standard errors for the main effect for block, indicates a fairly consistent decline in performance in blocks 1 - 6. Although there appears to be a slight upturn in tracking performance in blocks 5 and 6, post-hoc analyses did not show this to be significant. The LSD revealed that performance in block 1 was significantly better (lower RMSE) than in blocks 3 (p = .005), 4 (p = .000), 5 (p = .003), and 6 (p = .008). Performance in block 2 was also significantly better than in blocks 3 (p = .033), 4 (p = .002), 5 (p = .021) and 6 (p = .048). Blocks 3 - 6 did not differ significantly in tracking performance.

Table 13: Summary Analysis for CONLH Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	9838.87	14	5356.15	1.837	0.197
Cond	5	530.44	70	116.72	4.545	0.001
JTxCond	5	124.41	70	116.72	1.066	0.387

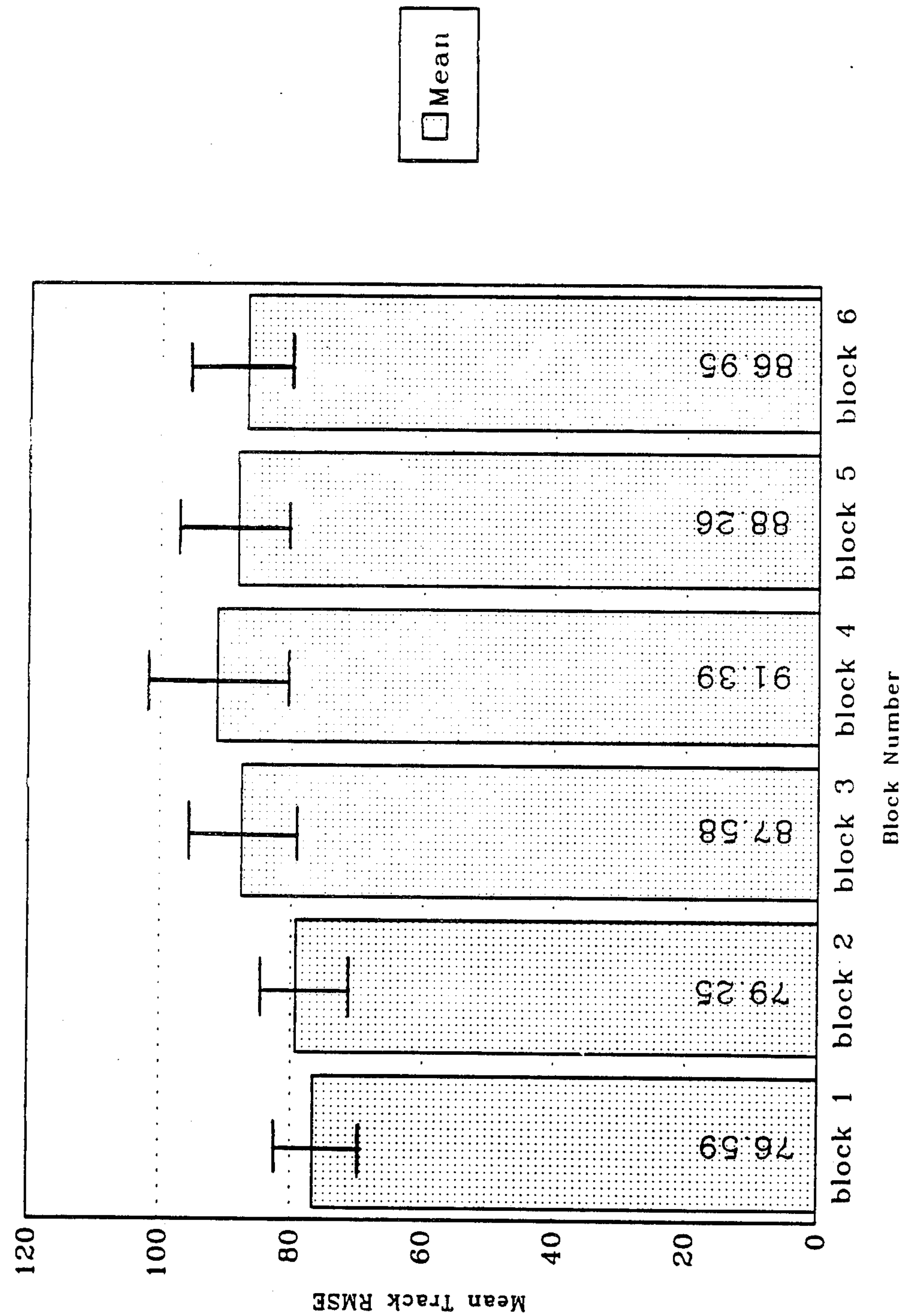


Figure 15. CONLH Block Means: Tracking RMSE Blocks 1-6

Considering the condition CONHL, Table 14 displays the summary analysis for tracking RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, **F** (5, 70) = 4.37, p < .01. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 2.40 and the degrees of freedom for error were corrected from 70 to 33.58. The p-value remained significant, correcting from p = .002 to p = .020.

Figure 16 displays the means and standard errors for the main effect for block. As shown in Figure 16, there is a fairly consistent decline in tracking performance occurring in blocks 1 - 6. Specifically, post-Hoc LSD analyses revealed that tracking performance in block 1 was significantly better (lower RMSE) than tracking performance in block 2 (p = 036), 3 (p = .005), 4 (p = .005), 5 (p = .016) and 6 (p = .000). Likewise, tracking performance in block 2 was significantly better than in block 6 (p = .018). Furthermore, there were no significant differences in tracking performance between blocks 3, 4, or 5. However, tracking performance in block 6 was significantly worse (higher RMSE) than the tracking performance in block 5 (p = .040).

Resource (Fuel) Management Task: Blocks 1-4

Again, during the first twenty minutes (Blocks 1-4) of each trial, either the system monitoring or the resource (fuel) management task was automated. The exception was in the case of the full manual controls. In the full manual control conditions, all three tasks were manually performed by the subjects for the entire session.

Table 14: Summary Analysis for CONHL Tracking RMSE Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	49.02	14	2119.84	0.023	0.881
Cond	5	1062.17	70	243.17	4.368	0.002
JTxCond	5	221.79	70	243.17	0.912	0.478

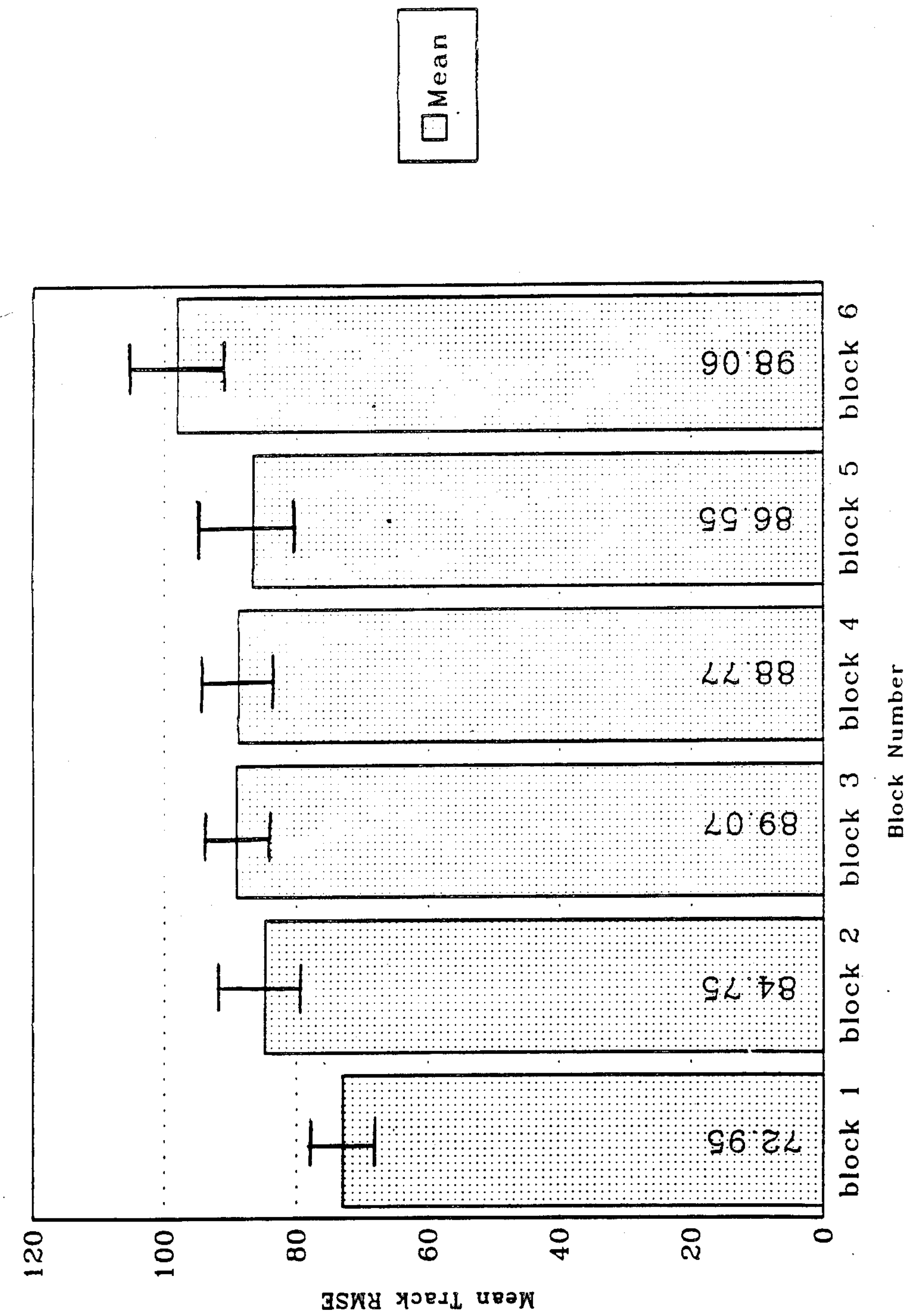


Figure 16: CONHIL Block Means
Tracking RMSE Blocks 1-6

A 3-Way Analysis of Variance (ANOVA) was performed on the dependent variable for resource (fuel) management, Root Mean Square Error (RMSE). The three factors entered into the analysis were judgment type (JT) (absolute vs. comparative), block number (Blk) (1, 2, 3, 4) and condition (Cond)(ASMHH, ASMHL, CONHH, CONLH, CONHL). Note that the conditions ARMHH and ARMLH were not entered into the analysis. These conditions were not entered because the resource (fuel) management task was automated during blocks 1 - 4 of these conditions. Therefore, blocks 1 - 4 of ARMHH and ARMLH contain no subject performance data on resource (fuel) management. The subjects were not manually performing the resource (fuel) management task at that time.

The purpose of this analysis was to determine the possible effects of automating the system monitoring task on performance within the resource (fuel) management task. Additionally, the full manual control conditions (CONHH, CONLH, CONHL) were examined to determine the possible effects of high versus low levels of engagement within both the system monitoring and the resource (fuel) management tasks on performance within the resource (fuel) management task.

Table 15 displays the summary analysis for the resource (fuel) management RMSE in blocks 1-4 across judgment types, block numbers, and conditions. As indicated in the summary table, there was a significant main effect for block number, $\underline{F}(3, 42) = 24.97$, $p < .01$. Additionally, there was a significant main effect for condition, $\underline{F}(4, 56) = 41.38$, $p < .01$. Finally, there was a significant two-way interaction between block number and condition, $\underline{F}(12, 168) = 34.74$, $p < .01$.

Table 15: Summary Analysis for Resource (Fuel) Managementt RMSE Blocks 1-4

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1688.16	14	27519.09	0.061	0.808
Blk	3	58005.60	42	2322.56	24.975	0.000
Cond	4	46746.96	56	1129.69	41.380	0.000
JTxBlk	3	2799.28	42	2322.56	1.205	0.319
JTxCond	4	2543.84	56	1129.69	2.252	0.749
BlkxCond	12	46648.86	168	1342.65	34.744	0.000
JTxBlkx Cond	12	1745.70	168	1342.65	1.300	0.222

The Greenhouse-Geisser procedure was applied to all significant effects. All remained significant. In the case of the main effect for block number, the degrees of freedom for effect adjusted from 3 to 2.17 and the degrees of freedom for error corrected form 42 to 30.44. The resulting p-value remained the same ($p = .0000$). In the case of the main effect for condition, the degrees of freedom for effect adjusted from 4 to 2.43 and the degrees of freedom for error adjusted from 56 to 34.04. Again, the p-value remained the same ($p = .0000$). Finally, in the case of the two-way interaction between block number and condition, the degrees of freedom for effect adjusted from 12 to 3.08 and the degrees of freedom for error adjusted from 168 to 43.16. Once more, the p-value remained the same ($p = .0000$). Figure 17 displays the means and standard errors for the main effect for block. The figure indicates a general decline in resource (fuel) management performance (increasing RMSE) through block 4, with the worst performance occurring in block 3.

LSD post-hoc analyses revealed that, in fact, resource (fuel) management performance in block 1 was significantly better than in blocks 2 ($p = .000$), 3 ($p = .000$) and 4 ($p = .000$). Likewise, performance in block 2 was significantly better than in block 3 ($p = .000$). However, performance in block 3 was significantly worse (higher RMSE) than in block 4 ($p = .000$).

As in the case of the tracking RMSE data, these findings generally follow what one would expect with a decline in vigilance over time, with the exception of block 4.

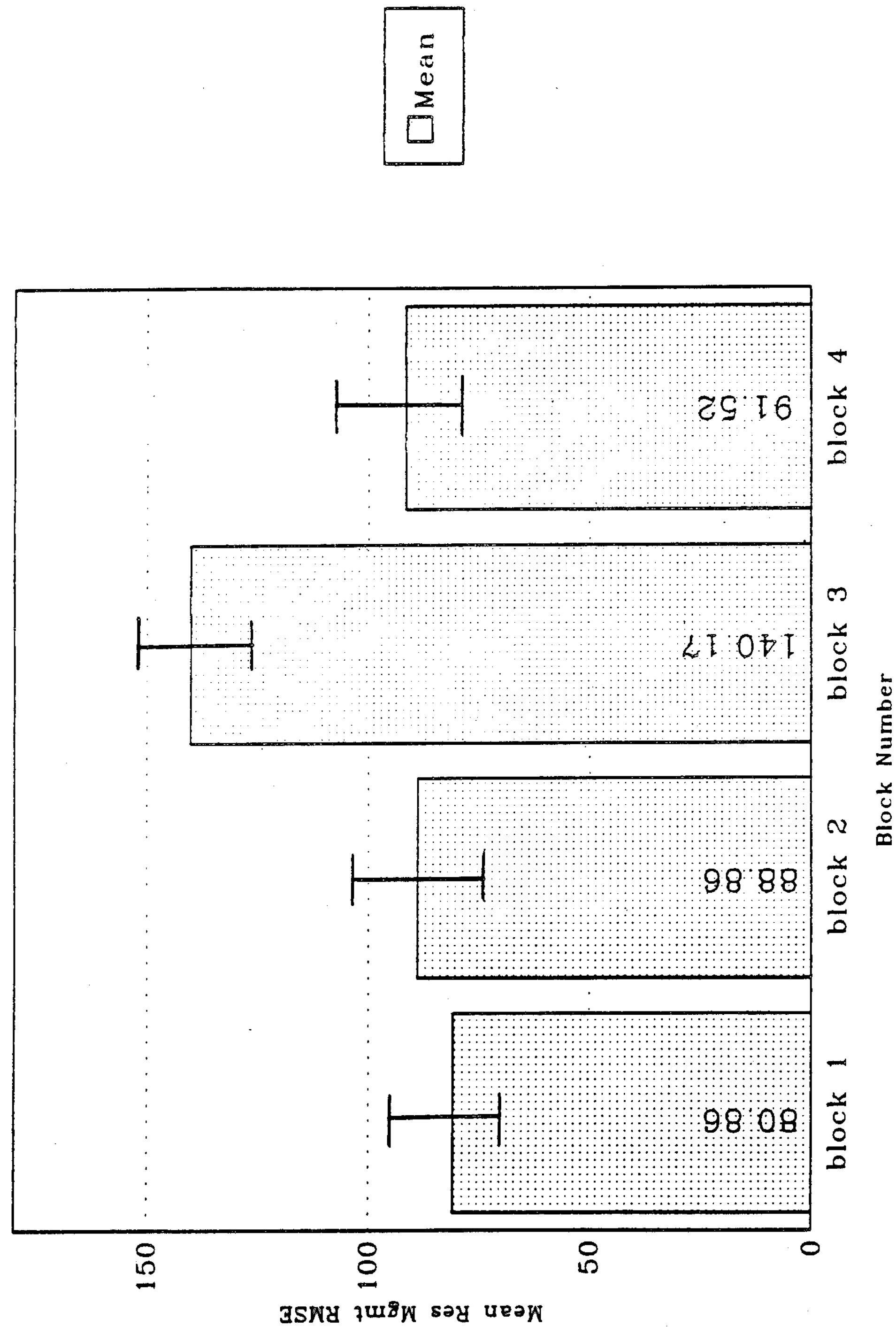


Figure 17. Block Means
Resource (Fuel) Management RMSE Blocks 1-4

The improvement in performance in block 4 may indicate subject ability to predict automation failure, despite the fact that subjects were told the automation would not necessarily fail each time.

Figure 18 displays the means and standard errors for the main effect for condition. The figure indicates a general decline in resource (fuel) management in the full manual control conditions, relative to those conditions in which the system monitoring task was automated. Post-Hoc LSD tests revealed that, as predicted, resource (fuel) management performance was aided by automation of the system monitoring task. In fact, performance in the condition ASMHH was significantly better (lower RMSE) than in its control, CONHH ($p = .000$) Additionally, performance in ASMHL was significantly better than in its control, CONHL ($p = .001$)

The post-hoc tests also revealed effects due to the level of engagement in both the system monitoring task and the resource (fuel) management task. Resource (fuel) management performance was significantly better under ASMHH than under ASMHL ($p = .010$). This indicates that, even though the system monitoring task was automated, a higher engagement level (6 signal per 5 minute block) in system monitoring aided performance in resource (fuel) management. Furthermore, resource (fuel) management performance was also significantly better under CONHH relative to CONLH ($p = .000$). This indicates that, as predicted, resource (fuel) management performance was better in blocks 1 - 4 when the engagement level of the resource management task was also high (6 pump failures per 5 minute block). According to Experiment I, Hypothesis # 2(c) (p. 30), a higher engagement level in the resource

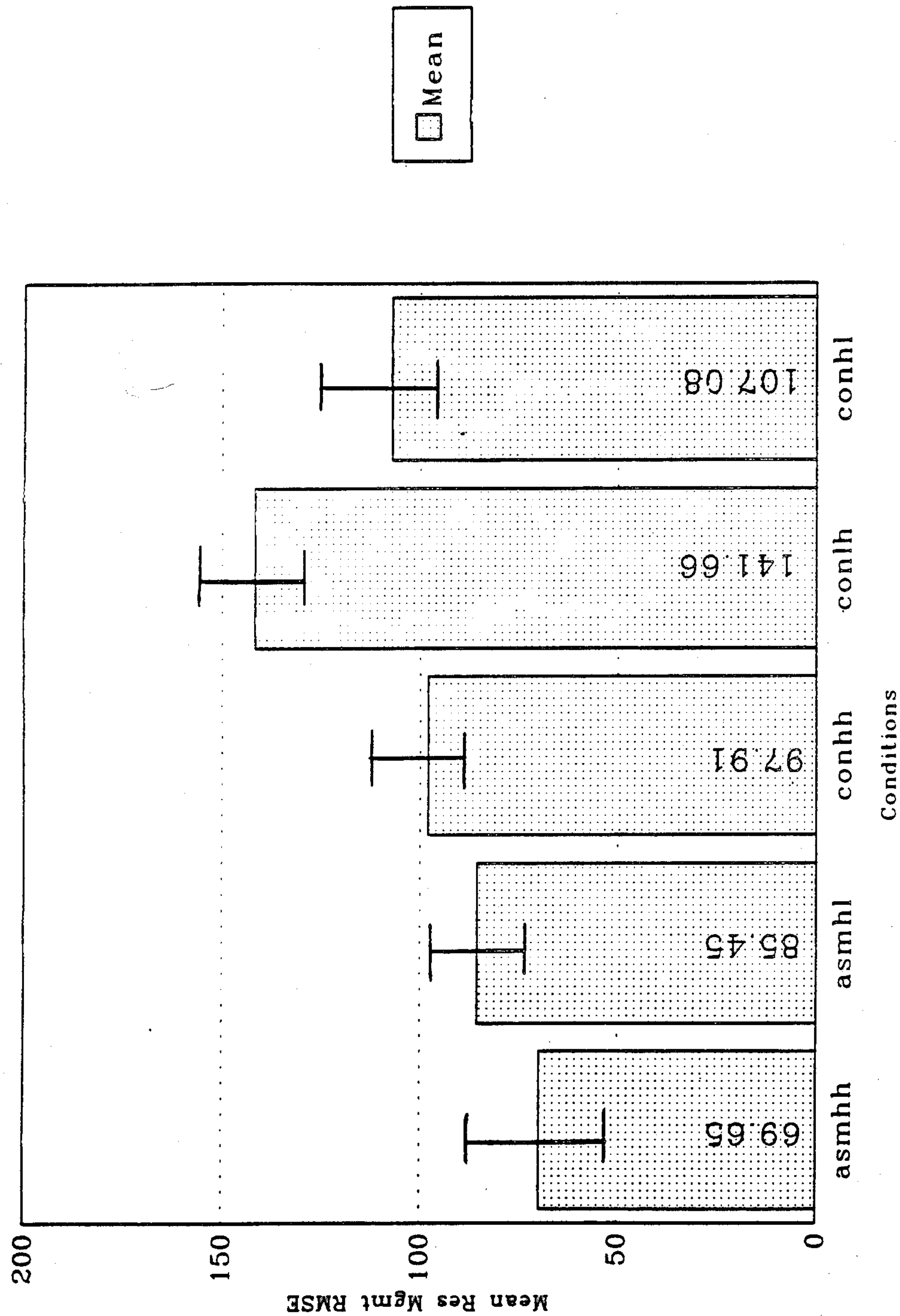


Figure 18: Condition Means
Resource (Fuel) Management RMSE Blocks 1-4

(fuel) management task was expected to increase subject involvement in the task, thereby allowing for increased updates in the dynamic model.

Figure 19 displays the means and standard errors for the two-way interaction between block number and condition. The figure indicates that, in general, within each of blocks 1 - 4, resource (fuel) management performance was aided by automation of the system monitoring task.

Post-Hoc LSD tests revealed that resource (fuel) management performance was significantly better while system monitoring was automated in all blocks except block 2. Specifically, in block 1 resource (fuel) management performance was significantly better (lower RMSE) under ASMHH as compared to both ASMHL ($p = .014$) and CONHH ($p = .000$). Likewise, performance was significantly better under ASMHL as compared to its control, CONHL ($p = .049$). However, performance in block 1 was significantly worse under CONHH as compared to both CONLH ($p = .058$) and CONHL ($p = .000$).

In block 2, the positive effects on resource (fuel) management performance due to automation of the system monitoring task were not seen. There were no significant differences between ASMHH and CONHH ($p = .678$) or between ASMHL and CONHL ($p = .114$).

In block 3, the positive effects on resource (fuel) management performance due to automation of the system monitoring task were seen. Performance was significantly better under ASMHL than under its control, CONHL ($p = .000$).

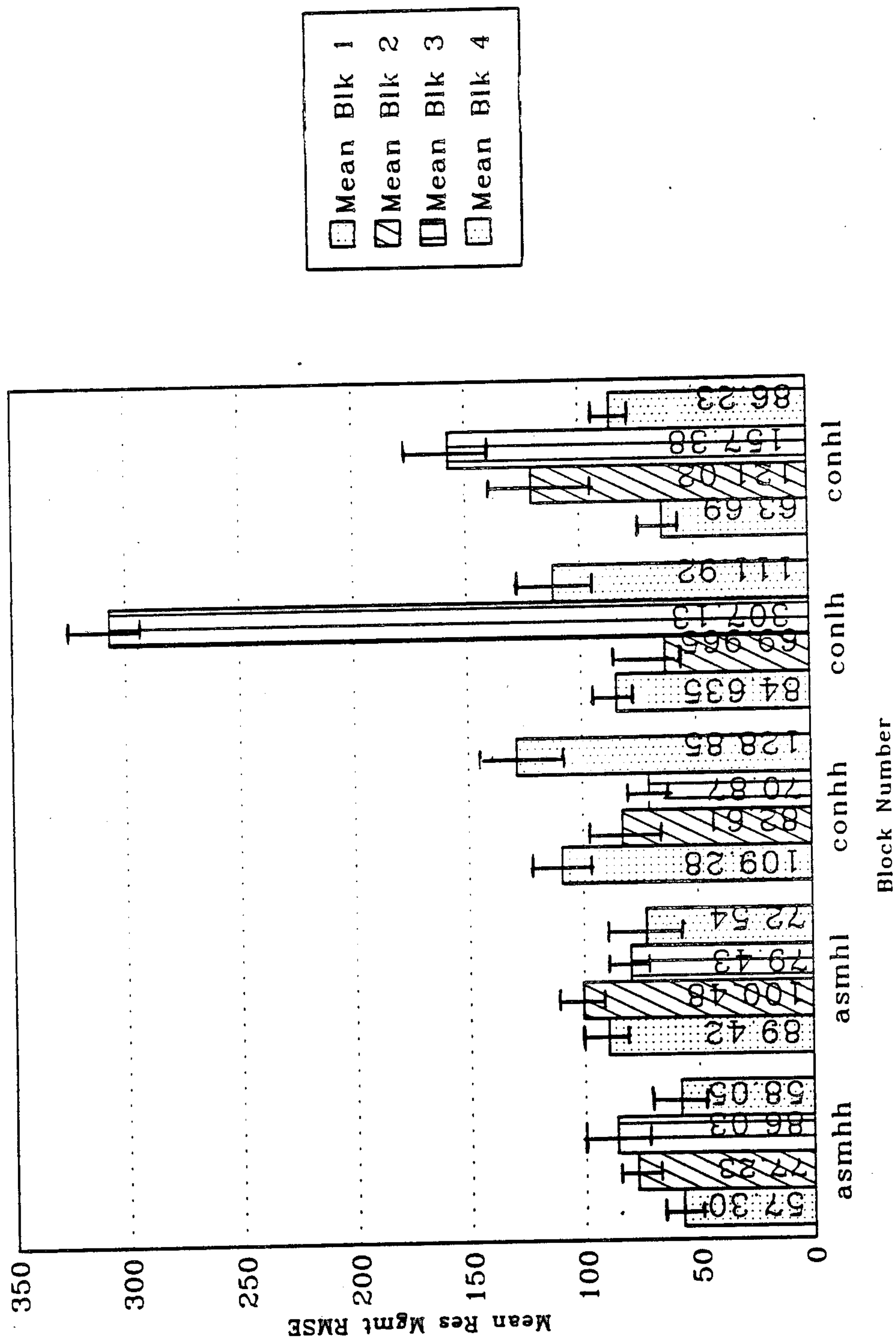


Figure 19: Block x Condition Interaction
Resource (Fuel) Management Blocks 1-4

In block 4, the positive effects on resource (fuel) management performance due to automation of the system monitoring task were seen. Performance under ASMHH was significantly better than under CONHH (p = .000).

These results support the expectation that, unlike in the stable model system monitoring task, performance in the more dynamic model resource (fuel) monitoring cannot be explained primarily in terms of vigilance. Furthermore, support was found for Experiment I, Hypothesis # 2 (c) (p. 30), which states that a higher level of engagement in the resource (fuel) management task should aid performance on that task prior to automation failure. Although there were no explicit hypotheses regarding effects of the engagement level of the system monitoring task on performance in the resource (fuel) management task, results seem to indicate improved performance in the resource (fuel) management task in the case of a higher engagement level in the system monitoring task, as well. This may indicate that the higher engagement levels in both tasks increase overall task vigilance (which is still a strong component in resource (fuel) management performance, just not as singularly as in system monitoring). An alternative explanation may be that the a discrepancy in the engagement levels of the two tasks (system monitoring and resource [fuel] management) was detrimental to performance.

Resource (Fuel) Management Task: Block 5

Block 5 of each session encompassed the first 5 minutes following automation failure of either the System Monitoring or the Resource (Fuel) Management task (except in the full manual control conditions, in which all three tasks were manually performed by the subject the entire session). Recall that subjects were instructed that upon detecting an automation failure, they were to regain manual control of the previously automated task, as well as to continue manual control of the other two tasks.

A 2-Way ANOVA was performed on resource (fuel) management RMSE for block 5. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was twofold. First, to determine the possible effects of long-term automation and automation failure in the system monitoring task on subsequent performance in the resource (fuel) management task. Second, and more importantly, to determine the possible effects of long-term automation and automation failure in the resource (fuel) management task on the subjects' performance in regaining manual control of that task.

Table 16 displays the summary analysis for the resource (fuel) management task RMSE in block 5 across judgment types and conditions. As indicated in the summary table, there was a significant main effect for condition, $\underline{\mathbf{F}}(6, 84) = 2.78$, $p < .05$. When the Greenhouse-Geisser was applied to the main effect for condition, the degrees of freedom for effect adjusted from 6 to 1.80 and the degrees of freedom for error adjusted from 84 to 25.25. The resulting p-level adjusted from $p = .016$ to $p = .081$, losing significance. However, the results were still interpreted with interest.

Table 16: Summary Analysis of Resource (Fuel) Management RMSE Block 5

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	34745.59	14	33782.00	1.028	0.328
Cond	6	20282.86	84	7288.04	2.783	0.016
JTxCond	6	10698.85	84	7288.04	1.468	0.199

An examination of Figure 20 shows decrements in performance in conditions where block 5 involves a regain in manual control of the resource (fuel) management task following automation of that task in blocks 1 - 4.

Post-hoc LSD analyses indicated that resource (fuel) management performance was significantly worse (higher RMSE) in block 5 of ARMHH than in block 5 of its control, CONHH ($p = .015$). Furthermore, resource (fuel) management performance was also significantly worse in block 5 of ARMHH than in block 5 of ASMHH ($p = .020$). In the case of ASMHH, block 5 still follows an automation failure, but of the system monitoring task. Thus, it indicates the decrement in resource (fuel) management resulted from attempts to regain manual control of that task, rather than as a result of automation failure, in general. This supports Experiment I, Hypothesis # 2(b) (p. 29). This hypothesis is that regaining manual control of a dynamic model task (such as resource (fuel) management), following long cycles of automation is more difficult, because the subject must first update the internal model in order to accurately assess and consequently diagnose the situation.

Resource (Fuel) Management Task: Block 6

Block 6 of each session encompassed the last 5 minutes of each session. A 2-Way ANOVA was performed on resource (fuel) management RMSE for block 6. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine any lingering effects of

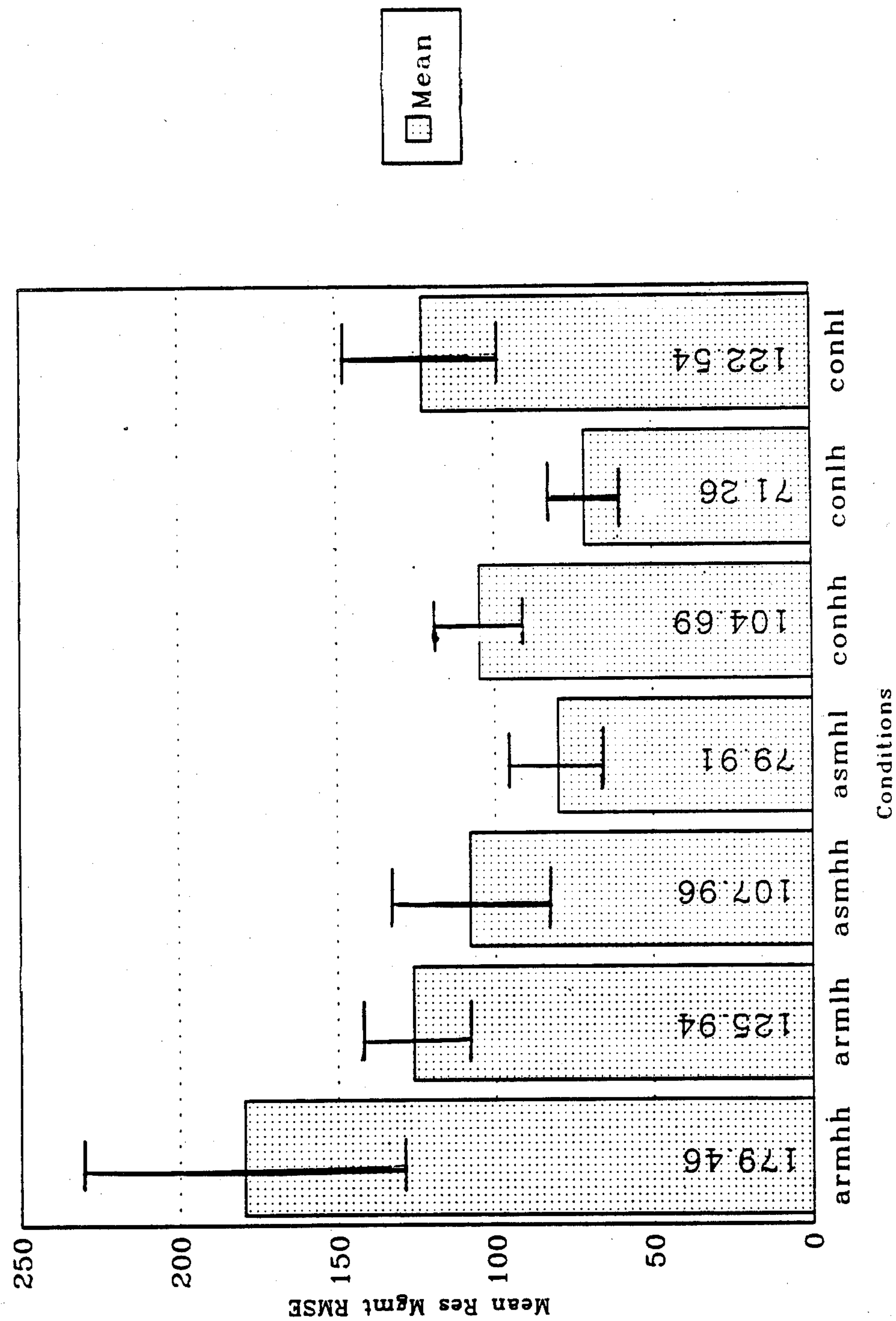


Figure 20: Condition Means :
Resource (Fuel) Management RMSE Block 5

automation and automation failure of the system monitoring task, and especially the resource (fuel) management task, on performance in the resource (fuel) management task.

Table 17 displays the summary analysis for the resource (fuel) management task RMSE in block 6 across judgment types and conditions.

As indicated in the summary table, there was a significant main effect for condition, $\underline{F}(6, 84) = 4.14$, $p < .01$. When the Greenhouse-Geisser was applied to the main effect for condition, the degrees of freedom for effect adjusted from 6 to 2.58 and the degrees of freedom for error adjusted from 84 to 35.39. The p-level remained significant, adjusting from $p = .001$ to $p = .013$.

An examination of Figure 21 shows decrements in performance into block 6 under conditions in which either resource (fuel) management or system monitoring was automated, but only when the engagement levels of both tasks were high. In other words, block 6 performance decrements are seen only under the conditions ARMHH and ASMHH. Post-Hoc LSD tests revealed that resource (fuel) management performance was significantly worse (higher RMSE) under ARMHH than under ARMLH ($p = .002$) and CONHH ($p = .007$). Additionally, performance under ASMHH was significantly worse than under ASMHL ($p = .003$) and CONHH ($p = .010$).

The fact that performance on the resource (fuel) management task was detrimentally affected in block 6 by ASMHH was not predicted. Because resource (fuel) management RMSE was lower in block 6 under ASMHH, as well as under

Table 17: Summary Analysis of Resource (Fuel) Management RMSE Block 6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	720.40	14	18706.22	0.038	0.847
Cond	6	6033.16	84	1457.82	4.138	0.001
JTxCond	6	2791.52	84	1457.82	1.915	0.087

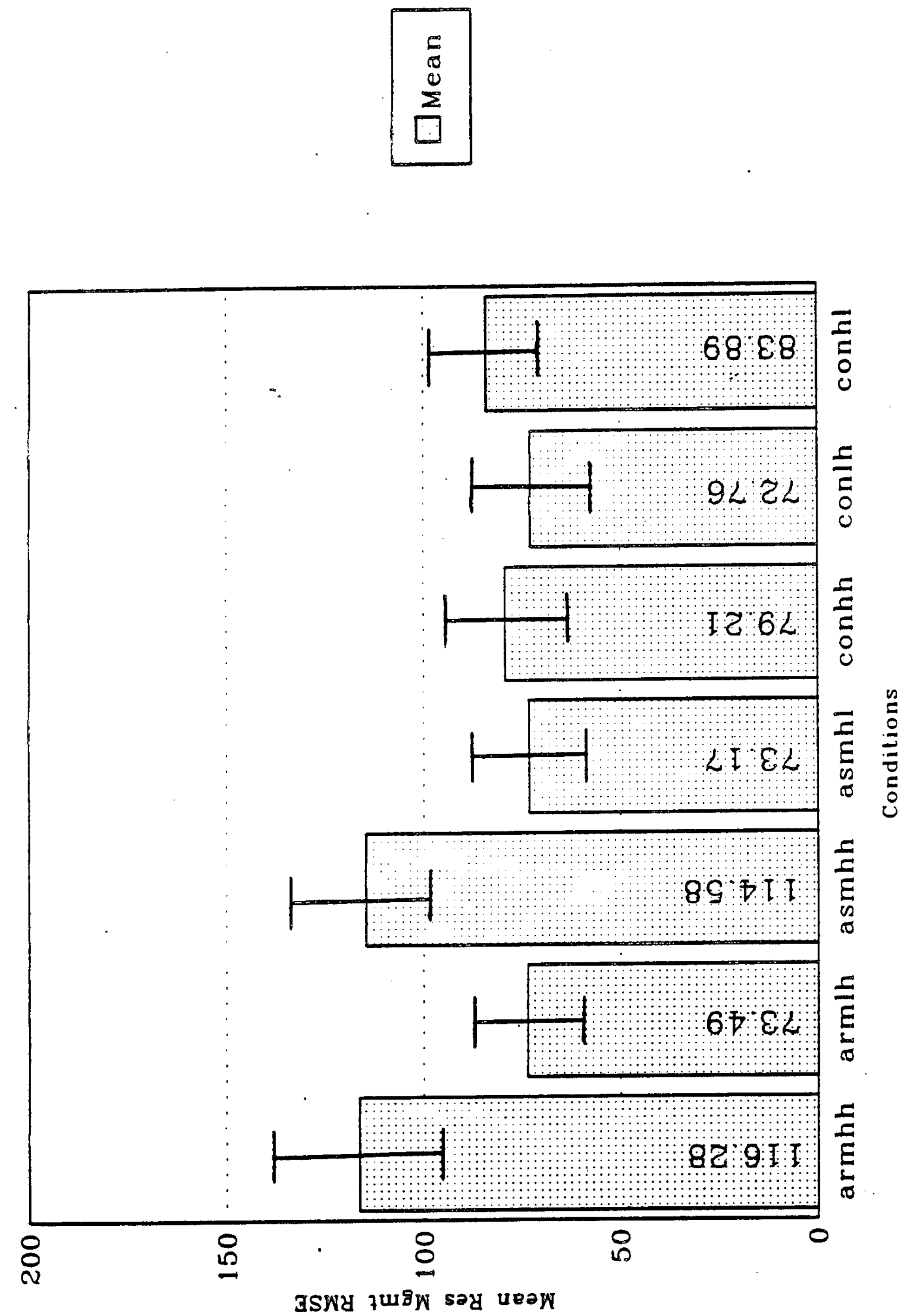


Figure 21. Condition Means
Resource (Fuel) Management RMSE Block 6

ARMHH, the performance detriments seen under ARMHH are probably not lingering effects of trying to regain manual control of the dynamic model task, resource (fuel) management, as was the case in block 5. The higher RMSE seen under both conditions ARMHH and ASMHH may be due to the combination of general automation failure and a high engagement level in both tasks affecting subjects in the final 5 minutes of the session.

Resource (Fuel) Management Task: Blocks 1-6

In order to examine any general trends in resource (fuel) management performance over an entire condition session, a separate 2-way ANOVA was performed on block 1-6 RMSE for each of the conditions, except for the two in which resource (fuel) management was automated during blocks 1-4 (ARMHH and ARMLH). These were not entered because no human resource (fuel) management performance data was gathered during blocks 1-4 of these conditions, as this task was automated at that time. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and block number (Blk) (1, 2, 3, 4, 5, 6).

Of the conditions examined, all had significant main effects for block. However, significance was lost in the condition ASMHL following the Greenhouse-Geisser adjustment. There was a general trend across the five conditions entered, in which resource (fuel) management performance fluctuated through block 6. Typically, performance was best in the initial blocks, fell slightly, rose again, and then fell again in the final blocks. The specific results of the separate analyses follow.

Considering the condition ASMHH, Table 18 displays the summary analysis for resource (fuel) management RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, **F** (5, 70) = 3.67, p < .01. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 2.11 and the degrees of freedom for error were corrected from 70 to 29.50. The p-value remained significant, with a correction from p = .005 to p = .038.

As can be seen in Figure 22, which displays the means and standard errors for the main effect for block, there is a fluctuation in resource (fuel) management performance over the 6 blocks. Performance starts off well, drops off consistently through block 3, rises in block 4, and drops off again through block 6. Post-hoc LSD tests revealed that performance in block 1 was significantly better (lower RMSE) than blocks 2 (p = .047), 3 (p = .002), 5 (p = .002), and 6 (p = .001).

Considering the condition ASMHL, Table 19 displays the summary analysis for resource (fuel) management RMSE in blocks 1-6. As indicated in the summary table, there was a significant main effect for block number, **F** (5, 70) = 2.44, p < .05.

With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 2.66 and the degrees of freedom for error were corrected from 70 to 37.18. The p-value lost significance, correcting from p = .043 to p = .080.

Though not significant, Figure 23 shows a fluctuation in resource (fuel) management performance over the 6 blocks. As in the condition ASMHH, performance starts off well, drops off consistently through block 3, rises in block 4, and begins to improve slightly in blocks 5 and 6.

Table 18: Summary Analysis for ASMHH Resource (Fuel) Management Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	12266.60	14	27519.09	0.889	0.362
Blk	5	7385.29	70	2322.56	3.666	0.005
JTxBlk	5	4636.17	70	1129.69	2.301	0.539

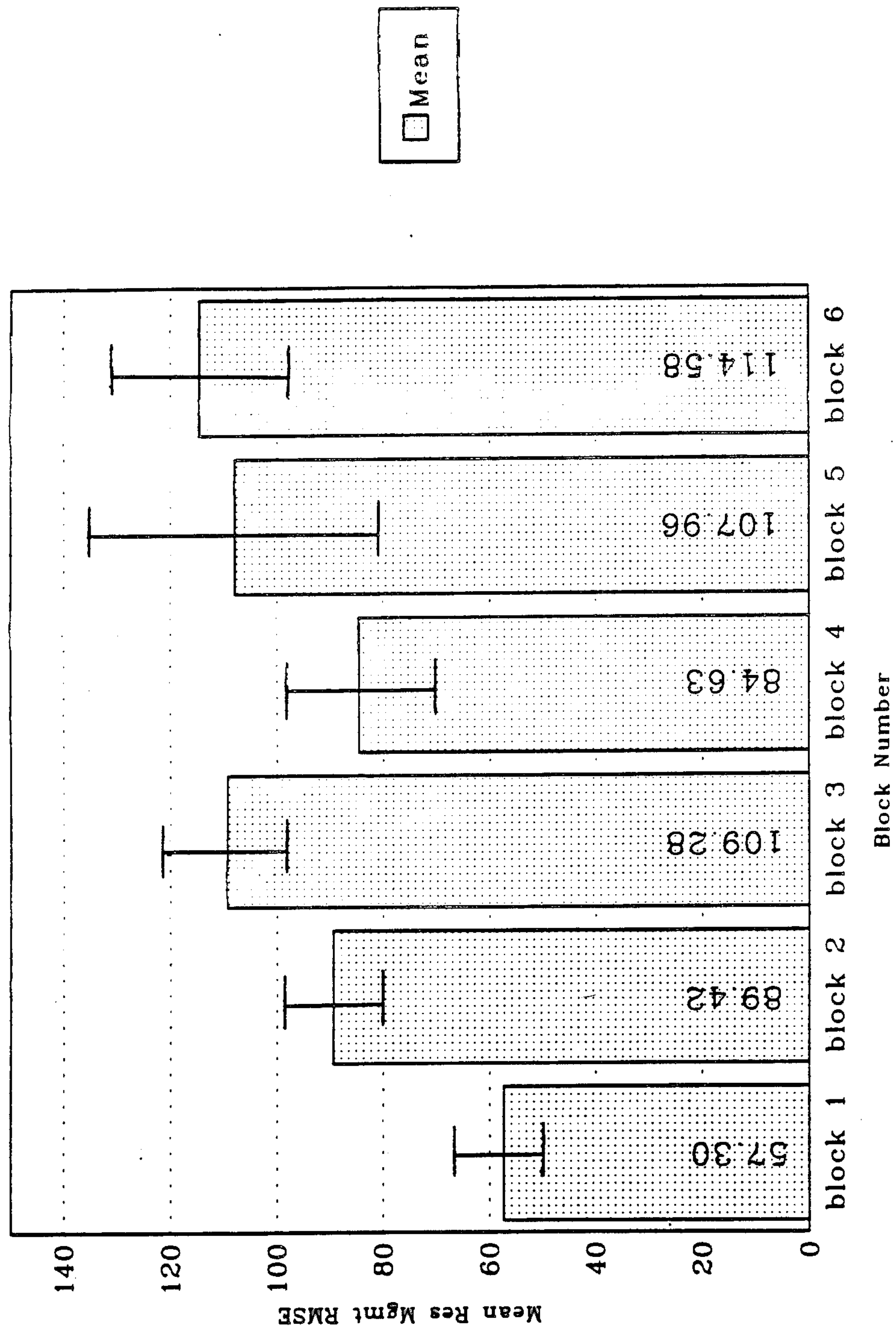


Figure 22. ASMHH Block Means
Resource (Fuel) Management RMSE Blocks 1-6

Table 19: Summary Analysis for ASMHL Resource (Fuel) Management Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	102.28	14	11497.49	0.009	0.926
Blk	5	2385.34	70	978.02	2.439	0.043
JTxBlk	5	439.81	70	978.02	0.449	0.812

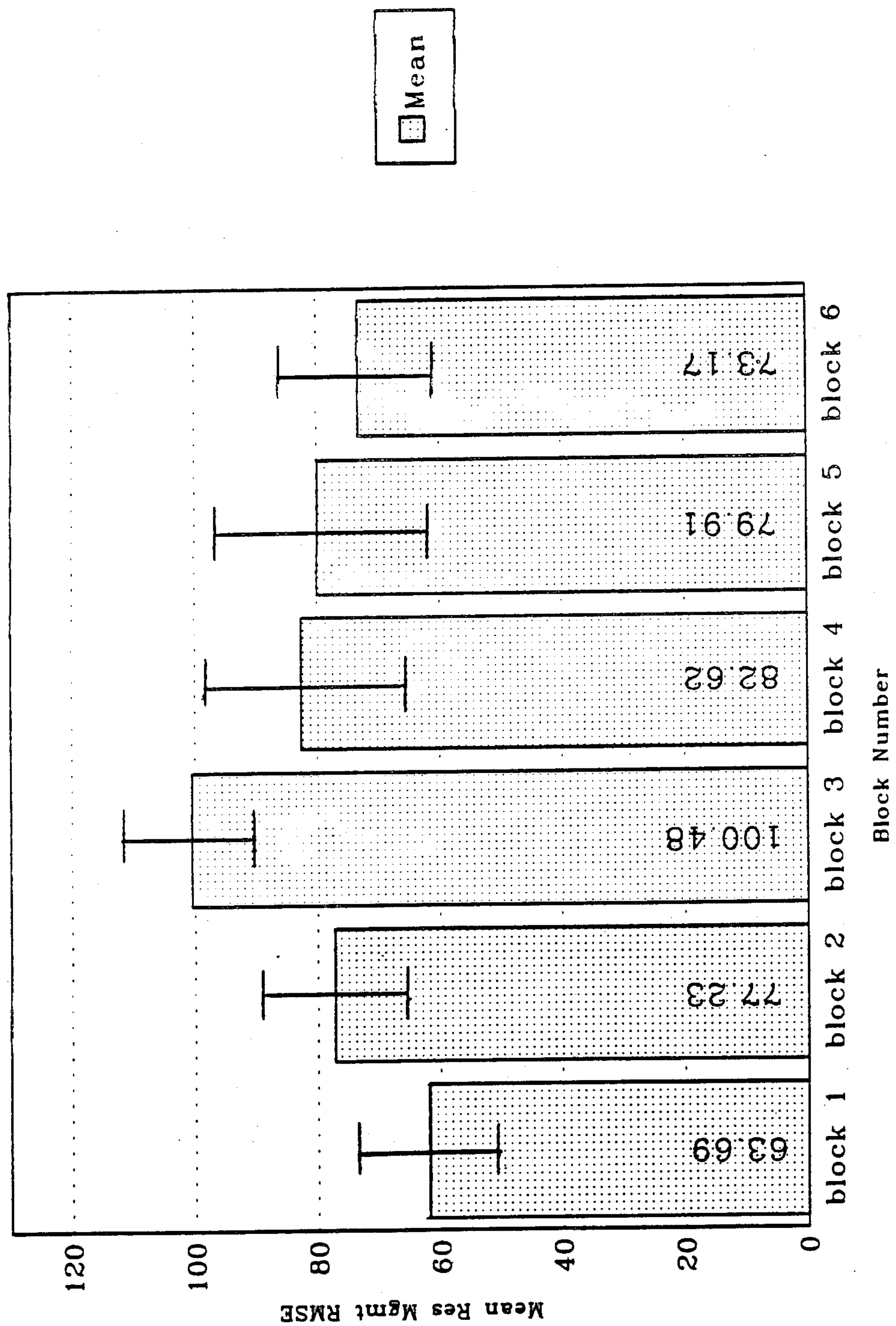


Figure 23. ASMIII, Block Means
Resource (Fuel) Management RMSE Blocks 1-6

Considering the condition CONHH, Table 20 displays the summary analysis for resource (fuel) management RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, **F** (5, 70) = 8.66, p < .01. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 3.06 and the degrees of freedom for error were corrected from 70 to 42.82. The p-value remained essentially the same at p = .000. There was also a significant two-way interaction between judgment type and block, **F** = 2.72, p < .05. However, this effect lost significance following the Greenhouse-Geisser adjustment, where degrees of freedom for effect adjusted from 5 to 3.06, degrees of freedom for error adjusted from 70 to 60.07, and the p-level adjusted from p = .026 to p = .057. As can be seen in Figure 24, which displays the means and standard errors for the main effect for block, there is a fluctuation in resource (fuel) management performance over the 6 blocks. Performance starts off well, drops off dramatically in block 2, rises again through block 4, drops in block 5, and finally, rises in block 6.

Post-hoc LSD tests revealed that performance in block 1 was significantly better (lower RMSE) than in blocks 2 (p = .000), 3 (p = .023) and 5 (p = .000). However, performance in block 2 was significantly worse (higher RMSE) than in blocks 3 (p = .001), 4 (p = .000) and 6 (p = .000). Performance in block 3 then improved and was significantly better than in block 5 (p = .065). Likewise, performance in block 4 was significantly better than in block 5 (p = .013). However, performance in block 5 was significantly worse than in block 6 (p = .013).

Table 20: Summary Analysis for CONHH Resource (Fuel) Management Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	39.501	14	11924.02	0.003	0.955
Blk	5	6867.18	70	193.39	8.656	0.000
JTxBlk	5	2456.91	70	193.39	2.719	0.027

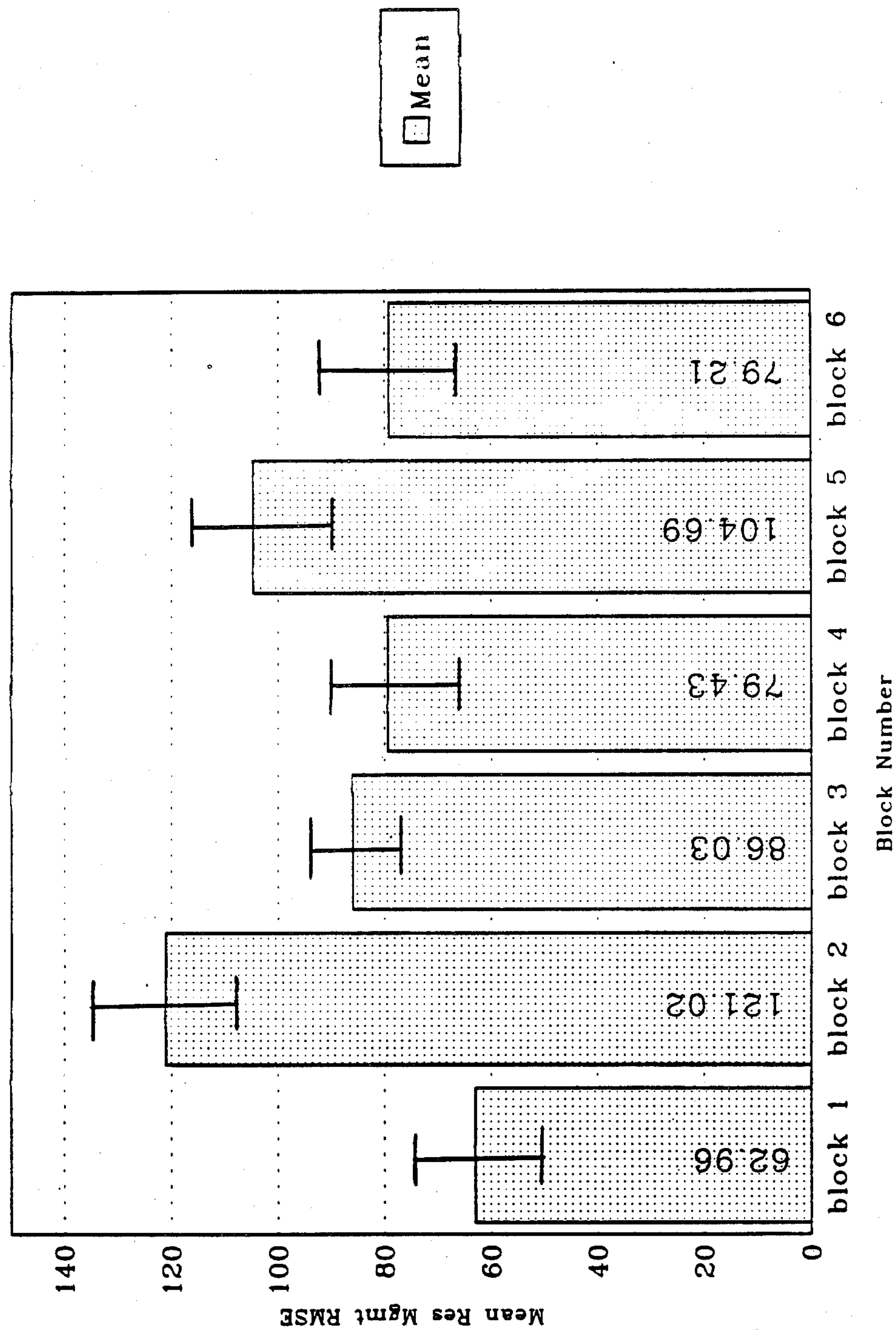


Figure 24. CONHIII Block Means
Resource (Fuel) Management RMSE Blocks 1 6

Considering the condition CONLH, Table 21 displays the summary analysis for resource (fuel) management RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, **F** $(5, 70) = 86.47$, $p < .01$. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were adjusted from 5 to 1.64 and the degrees of freedom for error were corrected from 70 to 22.98. The p-value remained significant and remained at $p = .000$. There was also a significant two-way interaction between judgment type and block, **F**$(5, 70) = 2.91$, $p < .05$. However, with the Greenhouse-Geisser adjustment, significance was lost. The degrees of freedom for effect were adjusted from 5 to 1.64 and the degrees of freedom for error were adjusted from 70 to 22.98, with a p-level adjustment from $p = .019$ to $p = .075$. As can be seen in Figure 25, which displays the means and standard errors for the main effect for block, there is a fluctuation in resource (fuel) management performance over the 6 blocks. Performance starts off well, drops off through block 2, rises through block 4, and drops off again in blocks 5 and 6. Post-hoc LSD tests revealed that performance in block 1 was significantly better (lower RMSE) than in blocks 2 ($p = .000$) and 3 ($p = .000$). Performance in block 2 then dropped sharply and was significantly worse than in blocks 3 ($p = .000$), 4 ($p = .000$), 5 ($p = .000$) and 6 ($p = .000$). Although performance in block 3 began to rise with respect to block 2, it was still significantly worse than in blocks 4 ($p = .000$), 5 ($p = .000$) and 6 ($p = .000$).

Table 21: Summary Analysis for COMLH Resource (Fuel) Management Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	4052.4	14	6969.49	0.581	0.458
Blk	5	151113.4	70	1747.59	86.469	0.000
JTxBlk	5	5086.5	70	1747.59	2.911	0.019

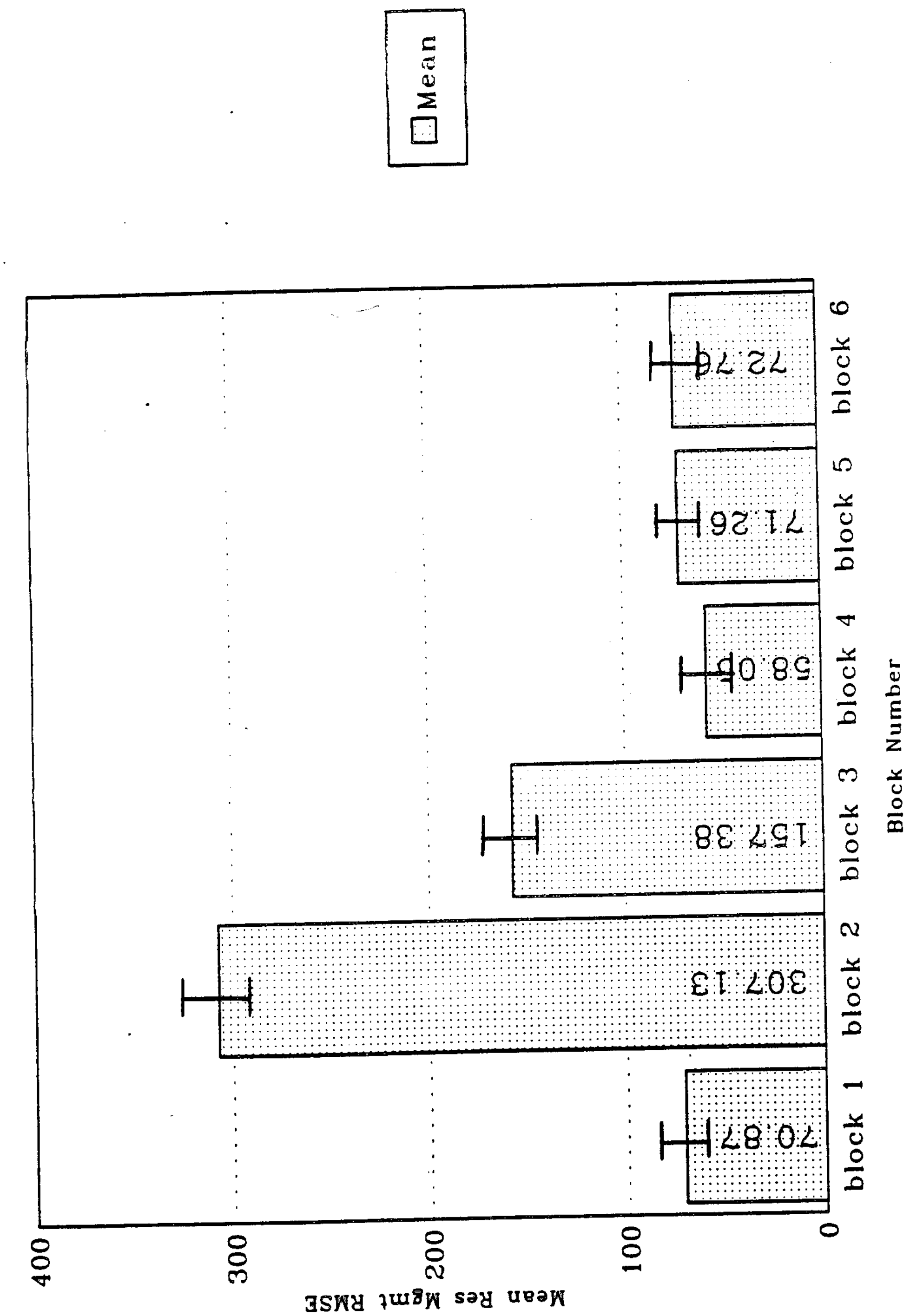


Figure 25: CONLII Block Means
Resource (Fuel) Management RMSE Blocks 1-6

Considering the condition CONHL, Table 22 displays the summary analysis for resource (fuel) management RMSE in blocks 1-6. As indicated in the summary table, there was a main effect for block number, **F** $(5, 70) = 3.31$, $p < .01$. With the Greenhouse-Geisser procedure, the degrees of freedom for effect were corrected from 5 to 1.94 and the degrees of freedom for error were corrected from 70 to 27.18. The p-value remained marginally significant, with a correction from $p = .009$ to $p = .052$.

As can be seen in Figure 26, which displays the means and standard errors for the main effect for block, there is a fluctuation in resource (fuel) management performance over the 6 blocks. Performance starts off well, drops off through block 3, rises in block 4, drops off again in block 5 and then rises again in block 6. Post-hoc LSD tests revealed that performance in block 1 was significantly better (lower RMSE) than blocks 2 ($p = .003$), 3 ($p = .032$) and 5 ($p = .007$). Performance in block 2 then drops and is significantly worse than in blocks 4 ($p = .021$) and 6 ($p = .015$). However, performance in block 4 improves and is significantly better than in block 5 ($p = .047$). Finally, performance in block 5 drops again and is significantly worse than in block 6 ($p = .035$).

System Monitoring Task: Reaction Time (RT) Blocks 1-4

Again, during the first twenty minutes (Blocks 1-4) of each trial, either the system monitoring or the resource (fuel) management task was automated, except in the case of the full manual controls, where all three tasks were manually performed by the subjects for the entire session. A 3-Way ANOVA was performed on the three dependent variables for the system monitoring task: median reaction time (RT), mean

Table 22: Summary Analysis for CONHL Resource (Fuel) Management Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1919.81	14	16641.05	0.115	0.739
Blk	5	8575.87	70	2590.14	3.311	0.009
JTxBlk	5	1712.09	70	2590.14	0.661	0.654

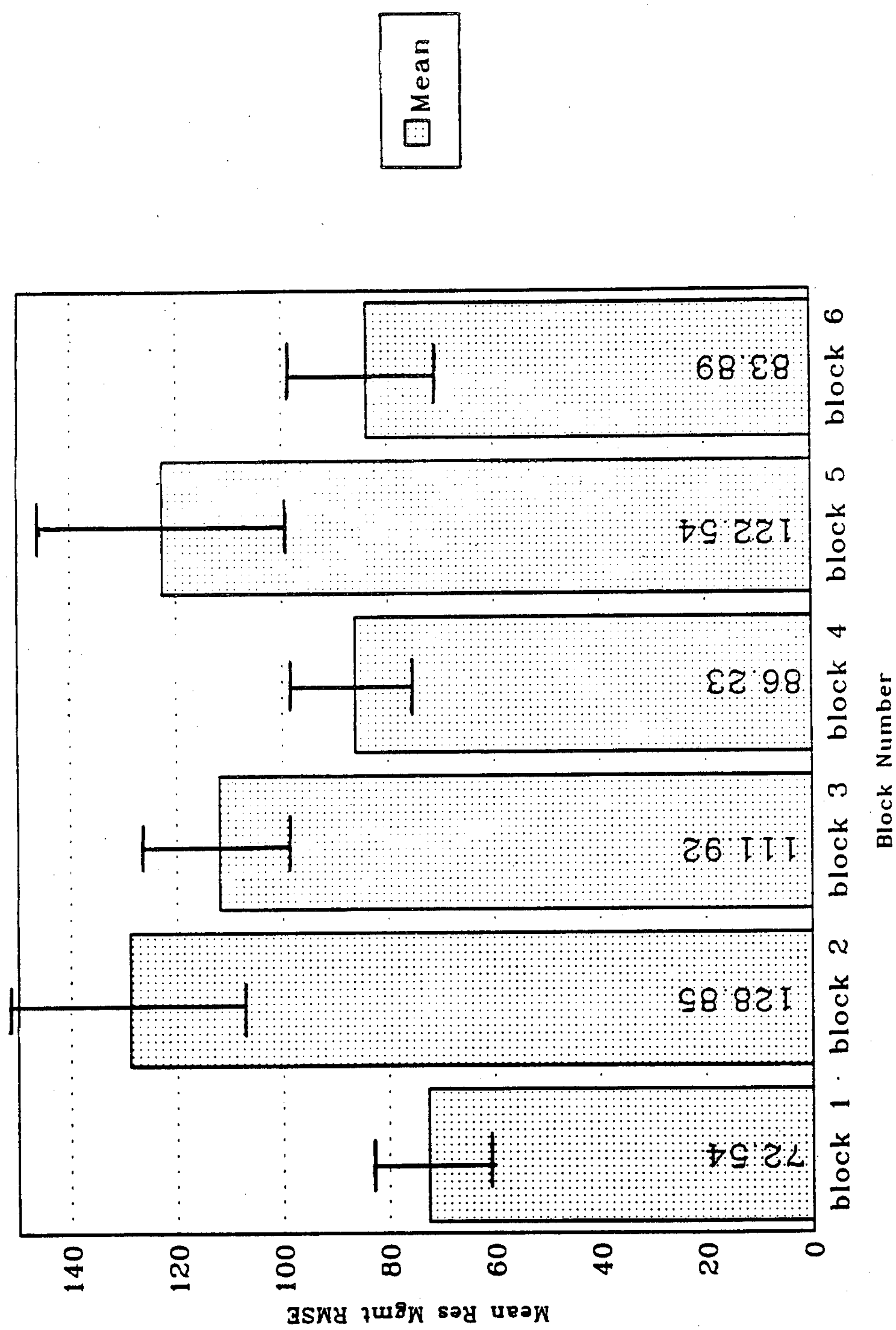


Figure 26. CONHIL Block Means
Resource (Fuel) Management RMSE Blocks 1-6

false alarm rate (FA) and mean percentage of correct detections (%CORR). The analysis presented in this section concerns the dependent variable median reaction time (RT). The three factors entered into the analysis were judgment type (JT) (absolute vs. comparative), block number (Blk) (1, 2, 3, 4) and condition (Cond)(ARMHH, ARMLH, CONHH, CONLH, CONHL). Note that the conditions ASMHH and ASMHL were not entered into the analysis. This is due to the fact that the system monitoring task was automated during blocks 1 - 4 of these conditions. Therefore, blocks 1 - 4 of ASMHH and ASMHL contain no subject performance data on system monitoring, as the subjects were not performing this task at that time.

The purpose of this analysis was to determine the possible effects of automating the resource (fuel) management task on reaction time performance within the system monitoring task. Additionally, the full manual control conditions were examined to determine the possible effects of high versus low levels of engagement within both the system monitoring and the resource (fuel) management tasks on performance within the system monitoring task.

Table 23 displays the Summary Analysis for the System Monitoring Median Reaction Time (RT) in blocks 1-4 across judgment types, block numbers, and conditions.

As indicated in the summary table, there were no significant main effects or interactions for this variable. As shown in Figure 27, which displays the means and standard errors for judgment type, median reaction time was somewhat slower under the absolute versus comparative judgment type. Again, this was not significant. Also

Table 23: Summary Analysis for System Monitoring Median RT Blocks 1-4

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	4.90	14	4.34	1.128	0.306
Blk	3	1.72	42	4.01	0.429	0.733
Cond	4	1.49	56	2.75	0.542	0.706
JTxBlk	3	2.60	42	4.01	0.648	0.588
JTxCond	4	5.10	56	2.75	1.857	0.131
BlkxCond	12	3.08	168	2.47	1.246	0.255
JTxBlkx Cond	12	2.28	168	2.47	0.924	0.524

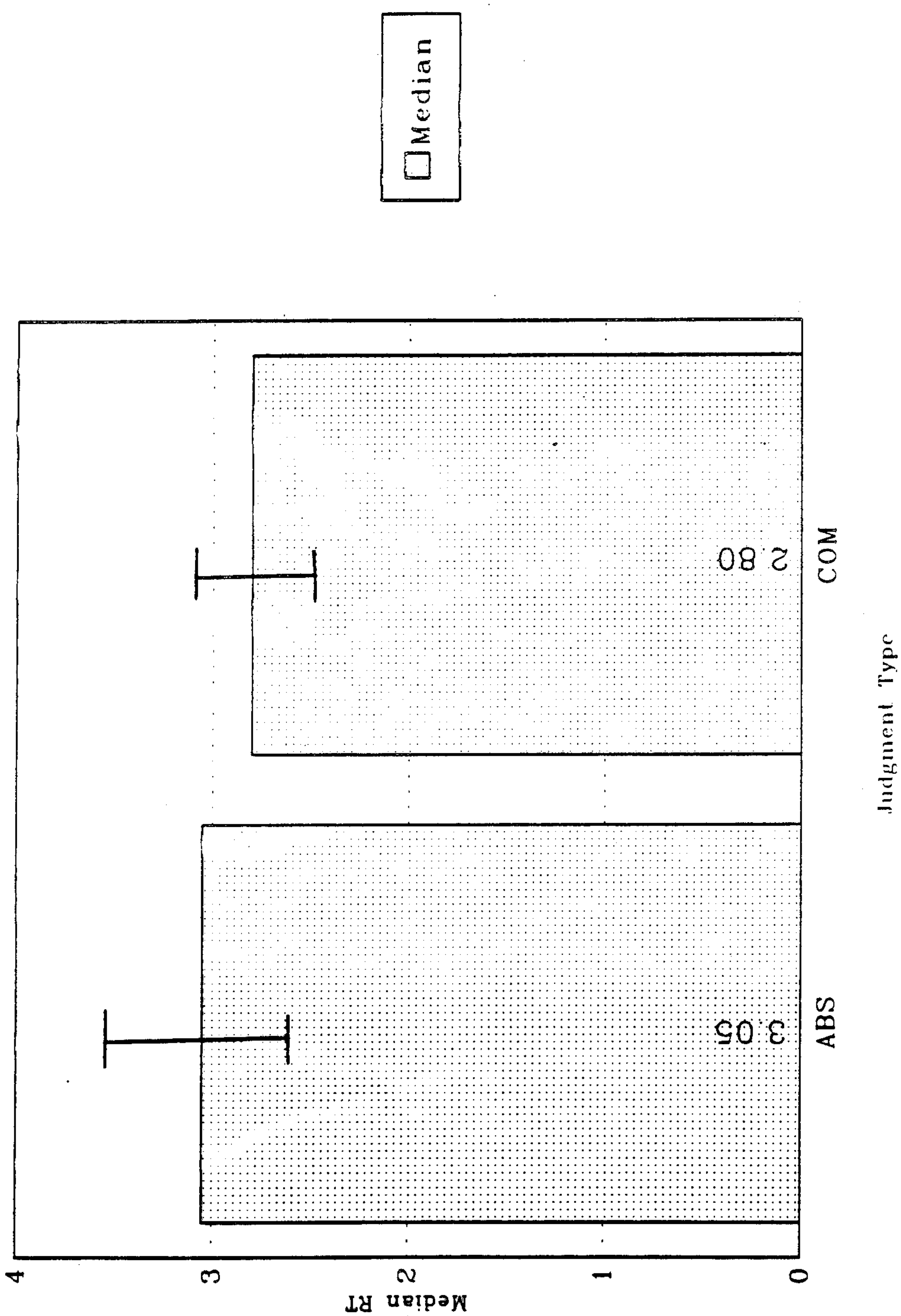


Figure 27 Judgment Type Medians
System Monitoring Reaction Time Blocks 1-4

of note is Figure 28. This figure presents the means and standard errors for block number. Figure 28 indicates a very slight but not significant increase in median reaction time over blocks.

System Monitoring Task: Reaction Time (RT) Block 5

Again, block 5 of each session encompassed the first 5 minutes following automation failure of either the system monitoring or the resource (fuel) management task (except in the full manual control conditions, in which all three tasks were manually performed by the subject the entire session). Recall that subjects were instructed that upon detecting an automation failure, they were to regain manual control of the previously automated task, as well as to continue manual control of the other two tasks.

A two-way ANOVA was performed on the dependent variable median reaction time for block 5. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was twofold.

First, to determine the possible effects of long-term automation and automation failure in the resource (fuel) management task on subsequent reaction time performance in the system monitoring task. Second, and more importantly, to determine the possible effects of long-term automation and automation failure in the system monitoring task on the subjects' performance in regaining manual control of that task.

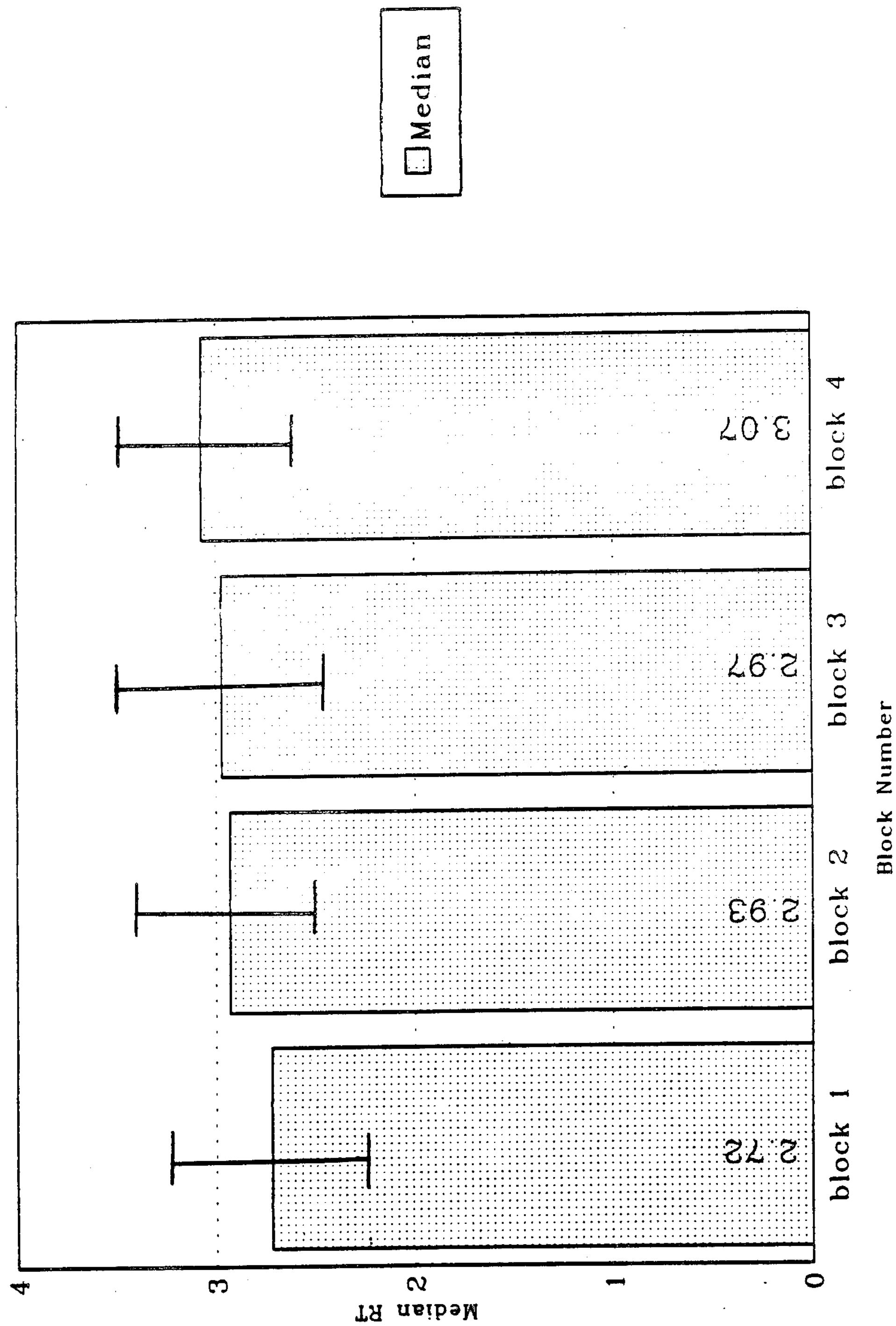


Figure 28. Block Medians
System Monitoring Reaction Time Blocks 1-4

Table 24 displays the summary analysis for the system monitoring task median reaction time in block 5 across judgment types and conditions.

As indicated in the summary table, there were no significant main effects or interactions in this analysis. However, as shown in Figure 29, there are medians of note under the conditions in which system monitoring was automated during blocks 1-4. Specifically, under the condition ASMHH, the median reaction time is somewhat shorter than the other conditions, whereas under the condition ASMHL, the median reaction time is somewhat longer than the other conditions. Though not significant, the pattern indicates some subject awareness of the system monitoring task while it was automated and would indicate benefits from higher a engagement level in regaining manual control of the system following automation failure, just as the higher engagement level benefits performance over long periods of time. This follows Experiment I, Hypothesis # 2(c) (p. 30), which specifies that a higher engagement level in the system monitoring task should benefit performance in both pre- and post-failure blocks.

System Monitoring Task: Reaction Time (RT) Block 6

Block 6 of each session encompassed the last 5 minutes of each session. A two-way ANOVA was performed on the dependent variable median reaction time for block 6. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine any lingering

Table 24: Summary Analysis for System Monitoring Median RT Block 5

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	7.58	14	5.04	1.506	0.240
Blk	6	7.97	84	4.06	1.960	0.805
JTxBlk	6	1.84	84	4.06	0.453	0.841

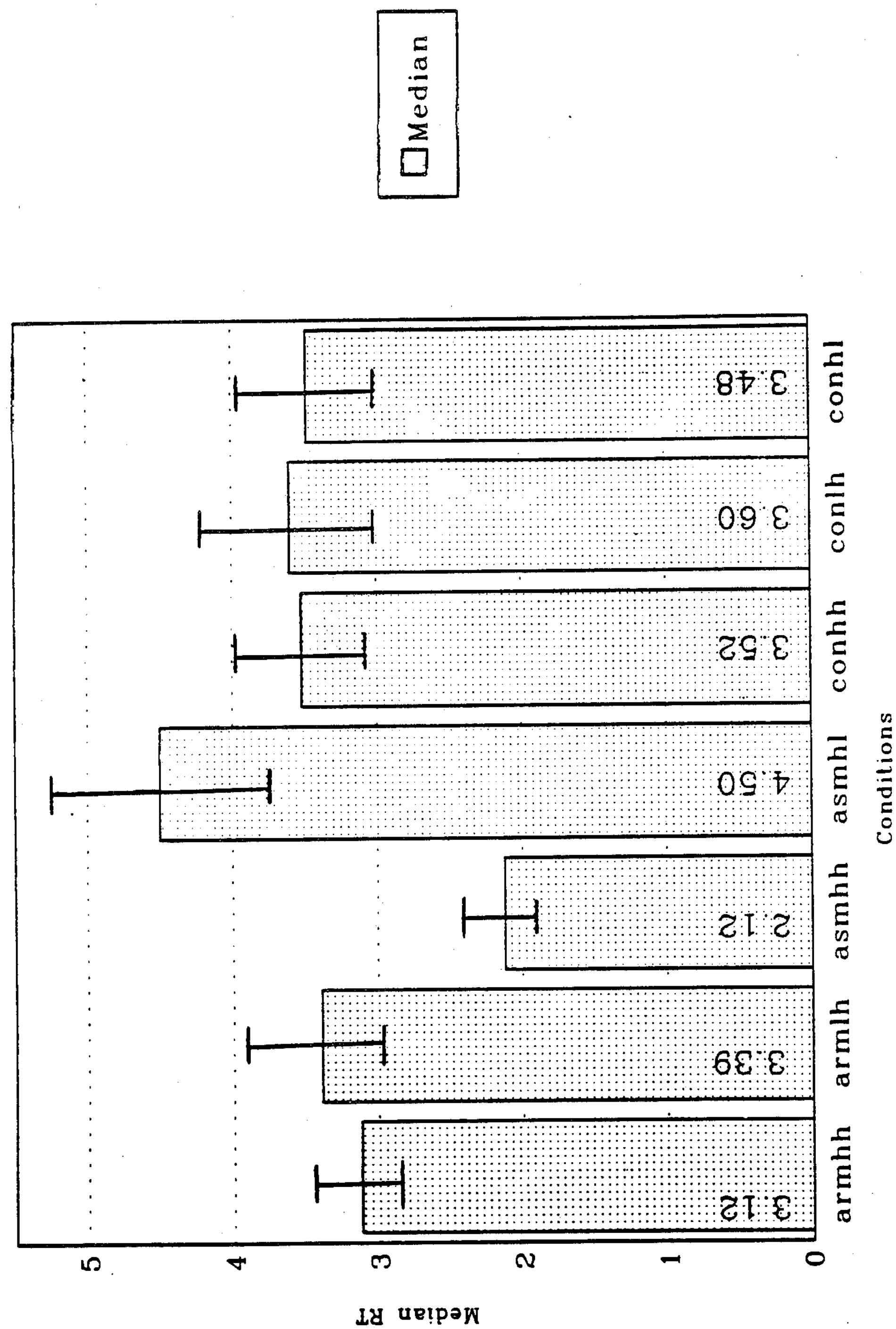


Figure 29: Condition Medians.
System Monitoring Reaction Time Block 5

effects of automation and automation failure of the resource (fuel) management task, and especially the system monitoring task, on reaction time performance in the system monitoring task.

Table 25 displays the summary analysis for the system monitoring task median reaction time in block 6 across judgment types and conditions.

As indicated in the summary table, there was a significant main effect for judgment type, $F(1, 14) = 5.09$, $p < .05$. An examination of Figure 30 shows decrements in performance (longer reaction time) into block 6 under the absolute judgment type, as predicted in Experiment I, Hypotheses # 3(a) (p. 32) and 4 (p. 32). The fact that this effect continued into block 6 indicates lasting effects for absolute versus comparative judgment on median reaction time.

System Monitoring Task: Reaction Time (RT) Blocks 1-6

In order to examine any general trends in reaction time over an entire condition session, a separate 2-way ANOVA was performed on block 1-6 median reaction times for each of the conditions, except for the two in which system monitoring was automated during blocks 1-4 (ASMHH and ASMHL). These were not entered because no human system monitoring performance data was gathered during blocks 1-4 of these conditions, as this task was automated at that time. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and block number (Blk) (1, 2, 3, 4, 5, 6).

Table 25: Summary Analysis for System Monitoring Median RT Block 6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	22.86	14	4.49	5.094	0.041
Blk	6	2.60	84	4.13	0.629	0.706
JTxBlk	6	4.21	84	4.13	1.019	0.418

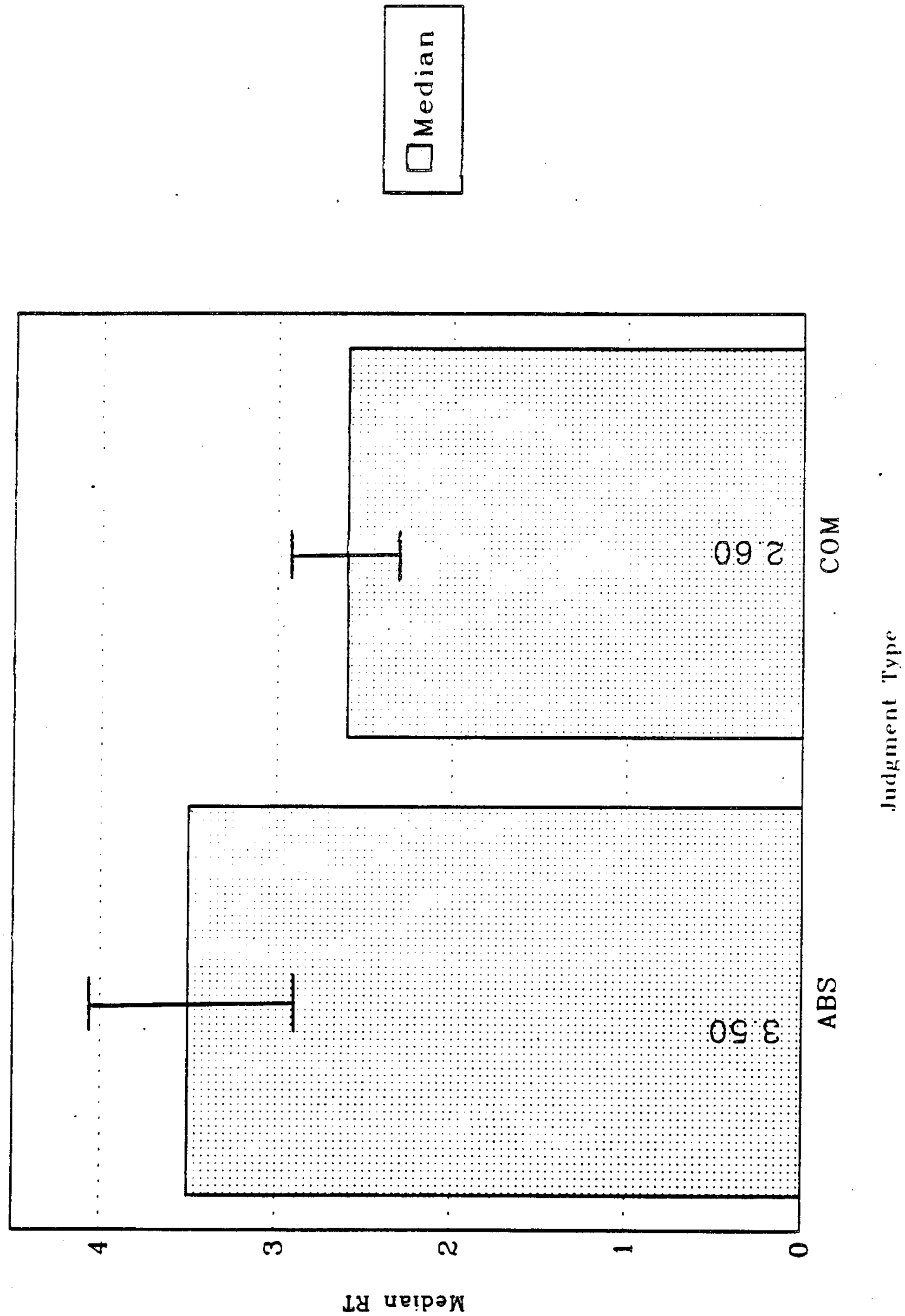


Figure 30. Judgment Type Medians
System Monitoring Reaction Time Block 6

Of the conditions examined, none had significant main effects for block, indicating there was not a significant change in reaction time over the 6 blocks. The summary analyses for the 5 conditions entered are displayed in Tables 26 through 30. The means and standard errors are presented in Appendix G.

System Monitoring Task: False Alarm Rate (FA) Blocks 1-4

A three-way ANOVA was performed on the dependent variable mean false alarm rate (FA). The three factors entered into the analysis were judgment type (JT) (absolute vs. comparative), block number (Blk) (1, 2, 3, 4) and condition (Cond)(ARMHH, ARMLH, CONHH, CONLH, CONHL). Again, the conditions ASMHH and ASMHL were not entered into the analysis because, during blocks 1-4, they did not contain human performance (i.e., subject) data on the system monitoring task.

The purpose of conducting this analysis was to determine the possible effects of automating the resource (fuel) management task on false alarm (FA) performance within the system monitoring task. Additionally, the full manual control conditions (CONHH, CONLH, CONHL) were examined. This was done in order to determine the possible effects of high levels of engagement versus low levels of engagement in the system monitoring task on performance within the system monitoring task, as well as to determine the possible effects of high levels of engagement versus low levels of engagement of the resource (fuel) management task on performance within the system monitoring task.

Table 26: Summary Analysis for ARMHH System Monitoring Median RT Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	9.93	14	4.12	2.409	0.143
Blk	5	3.03	70	1.54	1.966	0.094
Jtxblk	5	1.58	70	1.54	1.02	0.411

Table 27: Summary Analysis for ARMLH System Monitoring Median RT Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	5.33	14	4.26	1.253	0.282
Blk	5	2.44	70	2.23	1.094	0.372
JTxBlk	5	1.96	70	2.23	0.536	0.748

Table 28: Summary Analysis for CONHH System Monitoring Median RT Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	5.95	14	4.46	1.334	0.267
Blk	5	1.09	70	2.14	0.510	0.768
JTxBlk	5	3.99	70	2.14	1.867	0.111

Table 29: Summary Analysis for CONLH System Monitoring Median RT Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	19.73	14	6.07	3.250	0.093
Blk	5	4.00	70	3.97	1.006	0.421
JTxBlk	5	5.63	70	3.97	1.417	0.229

Table 30: Summary Analysis for CONHL System Monitoring Median RT Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.82	14	4.46	0.186	0.673
Blk	5	3.82	70	4.41	0.867	0.508
JTxBlk	5	2.93	70	4.41	0.663	0.652

Table 31 displays the summary analysis for the system monitoring mean false alarm rate (FA) in blocks 1-4 across judgment types, block numbers, and conditions. As indicated in the summary table, there was a significant main effect for block number, $\underline{F}(3, 42) = 3.14$, $p < .05$. However, when subjected to the Greenhouse-Geisser procedure, the degrees of freedom for effect corrected from 3 to 1.69 and the degrees of freedom for error corrected from 42 to 23.669. The resulting p-value was not significant, correcting from $p = .035$ to $p = .061$.

Though not significant, it is worth noting that the means and standard errors for judgment type, displayed in Figure 31 show a slight increase in false alarm rate under the absolute judgment type. This is in line with Experiment I, Hypotheses # 3(a) (p. 32) and # 4 (p. 32), which predicted worse performance under the absolute judgment type, particularly for the system monitoring task.

System Monitoring Task: False Alarm Rate (FA) Block 5

A two-way ANOVA was performed on the dependent variable mean false alarm rate for block 5. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was twofold.

First, to determine the possible effects of long-term automation and automation failure in the resource (fuel) management task on subsequent false alarm performance in the system monitoring task. Second, and more importantly, to determine the possible effects of long-term automation and automation failure in the system monitoring task on the subjects' performance in regaining manual control of that task.

Table 31: Summary Analysis for System Monitoring False Alarms Blocks 1-4

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	19.012	14	5.18	3.673	0.759
Blk	3	6.17	42	1.96	3.142	0.035
Cond	4	1.38	56	1.27	1.087	0.372
JTxBlk	3	2.67	42	1.96	1.360	0.268
JTxCond	4	0.35	56	1.27	0.274	0.893
BlkxCond	12	0.82	168	1.01	0.809	0.641
JTxBlkx Cond	12	0.99	168	0.01	0.983	0.467

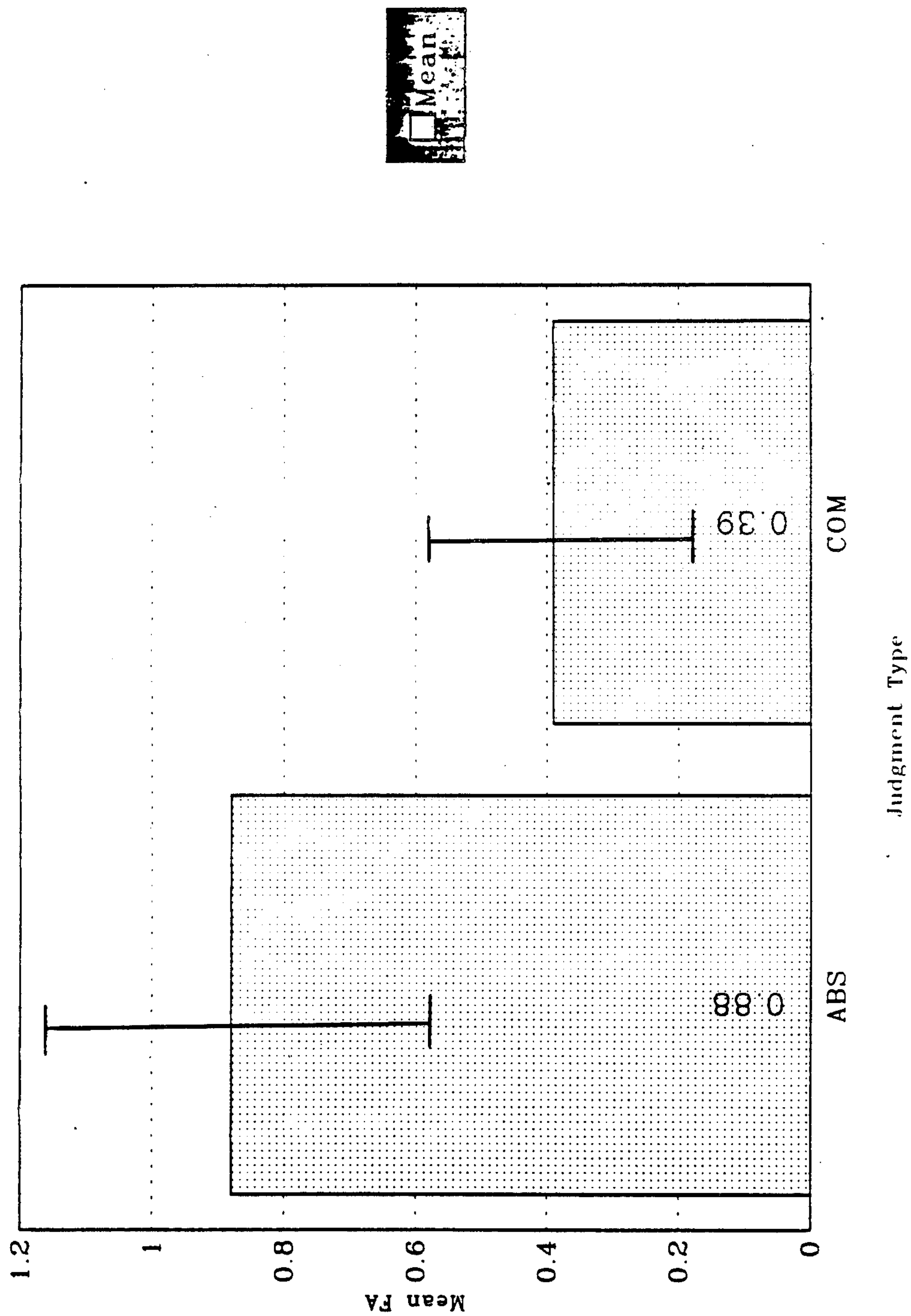


Figure 31. Judgment Type Means
System Monitoring False Alarms Blocks 1-4

Table 32 displays the summary analysis for the system monitoring task mean false alarm rate in block 5 across judgment types and conditions.

As indicated in the summary table, there was both a significant main effect for judgment type, **F** = 6.09, $p < .05$ and a significant interaction between judgment type and condition, **F** = 1.23, $p < .05$. The interaction, however, lost significance following the Greenhouse-Geisser procedure. The Greenhouse-Geisser adjusted the degrees of freedom for effect from 6 to 2.21 and the degrees of freedom for error from 84 to 30.93. The p-level adjusted from $p = .047$ to $p = .123$.

Regarding the significant main effect for judgment type, an examination of the means in Figure 32 shows that, as predicted in Experiment I, Hypotheses # 3(a) (p. 32) and # 4 (p. 32), the false alarm rate was significantly higher under the absolute judgment type.

System Monitoring Task: False Alarm Rate (FA) Block 6

A two-way ANOVA was performed on the dependent variable mean false alarm rate for block 6. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine any lingering effects of automation and automation failure of the resource (fuel) management task, and especially the system monitoring task, on false alarm rate in the system monitoring task.

Table 33 displays the summary analysis for the system monitoring task mean false alarm rate in block 6 across judgment types and conditions.

Table 32: Summary Analysis for System Monitoring False Alarms Block 5

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	7.51	14	1.23	6.088	0.027
Cond	6	1.97	84	1.23	1.599	0.158
JTxCond	6	2.76	84	1.23	2.242	0.047

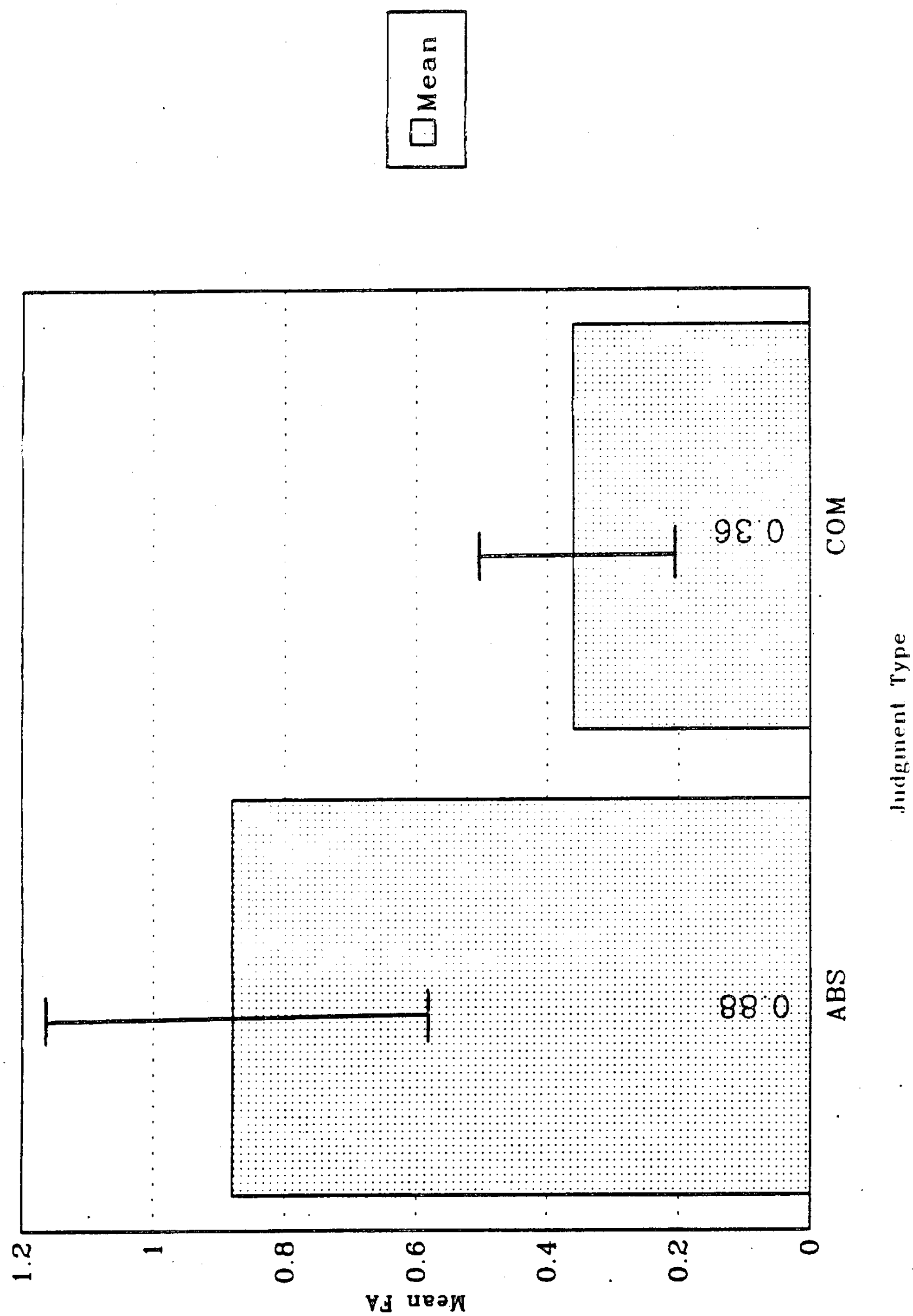


Figure 32. Judgment Type Means
System Monitoring False Alarms Block 5

Table 33: Summary Analysis for System Monitoring False Alarms Block 6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	15.75	14	2.99	5.259	0.038
Cond	6	1.39	84	1.17	1.184	0.322
JTxCond	6	0.56	84	1.17	0.479	0.822

As indicated in the summary table, there was a significant main effect for judgment type, **F** =5.26, $p < .05$. The means for the two judgment types, displayed in Figure 33, show that, as predicted, the false alarm rate was significantly higher under the absolute judgment type.

System Monitoring Task: False Alarm Rate (FA) Blocks 1-6

In order to examine any general trends in false alarm rate over an entire condition session, a separate 2-way ANOVA was performed on block 1-6 mean false alarm rate for each of the conditions, except for the two in which system monitoring was automated during blocks 1-4 (ASMHH and ASMHL). These were not entered because no human system monitoring performance data was gathered during blocks 1-4 of these conditions, as this task was automated at that time. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and block number (Blk) (1, 2, 3, 4, 5, 6). Of the conditions examined, only one had a significant main effect or interaction. Under CONHH, summarized in Table 34, there was a significant main effect for block, **F** = 3.07, $p < .05$. When subjected to the Greenhouse-Geisser procedure, the degrees of freedom for effect adjusted from 5 to 2.90 and the degrees of freedom for error adjusted from 70 to 40.63. The p-level remained significant, adjusting from $p = 0.014$ to $p = 0.038$. As shown in Figure 34, the false alarm rate increased through block 3, dropped somewhat in block 4, and moreso in block 5. Post-hoc LSD analysis indicated that only block 3 was significantly different from any other block. The false alarm rate in block 3 was

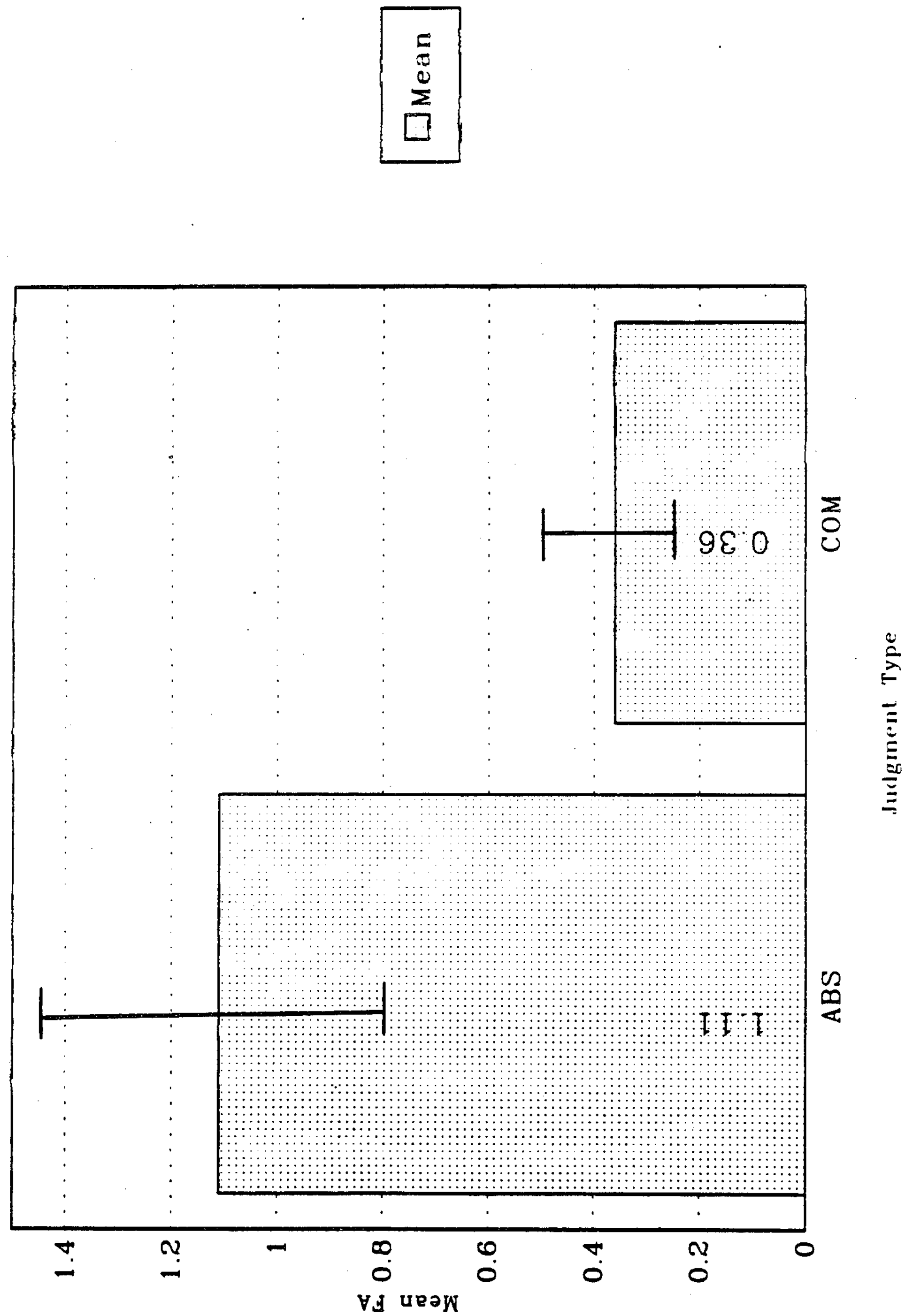


Figure 33: Judgment Type Means
System Monitoring False Alarms Block 6

Table 34: Summary Analysis for CONHH System Monitoring False Alarms Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	10.67	14	5.66	1.883	0.192
Blk	5	2.72	70	0.88	3.069	0.015
JTxBlk	5	1.09	70	0.88	1.233	0.303

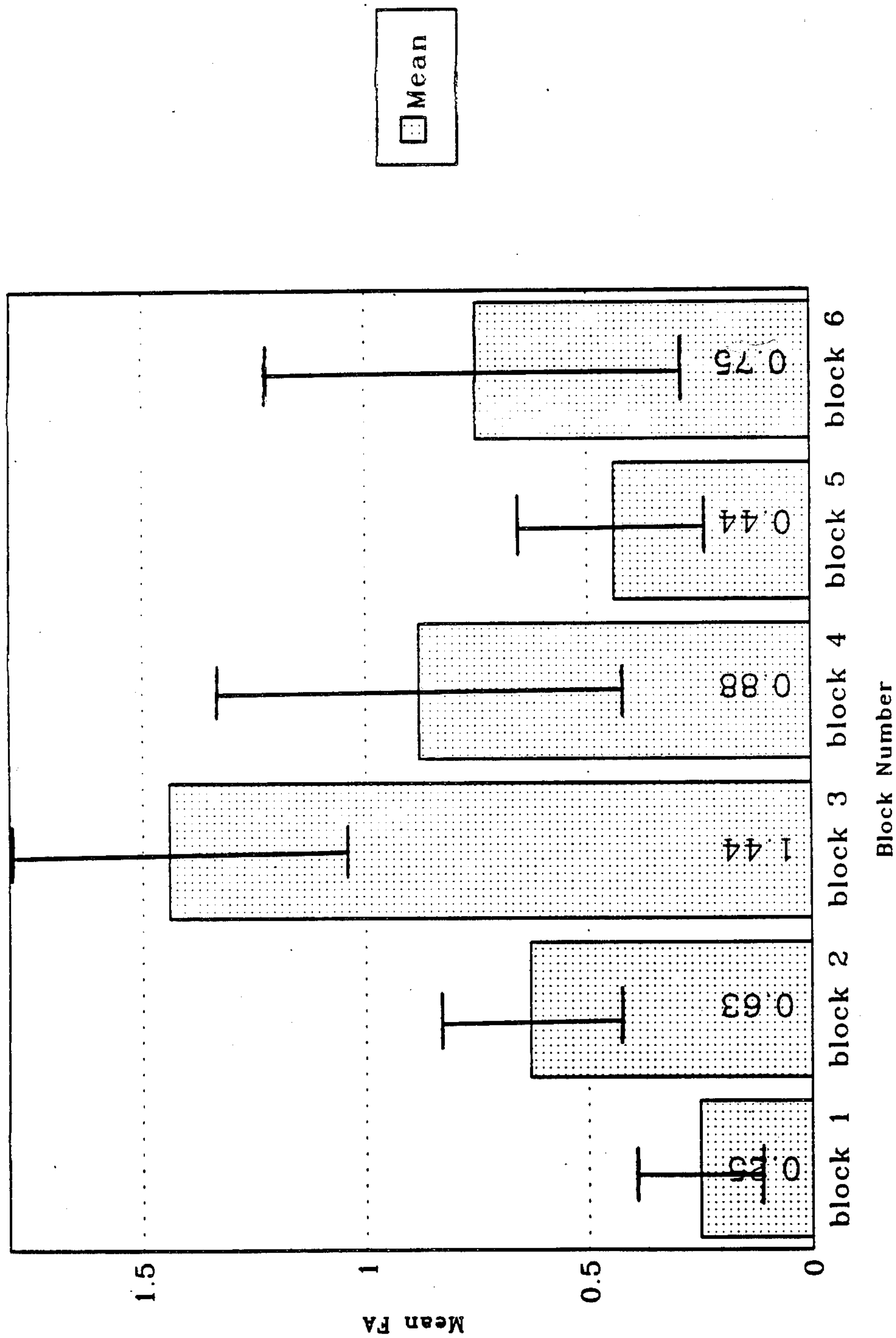


Figure 34: Block Means
System Monitoring Mean FA Blocks 1-6

significantly higher than in blocks 1 ($p = 0.001$), 2 ($p = 0.017$), 5 ($p = 0.004$) and 6 ($p = 0.042$). This seems to indicate a fluctuation in vigilance over time, with some improvement in block 3. However, note that performance in the later blocks never achieves the levels of blocks 1 and 2, and, in fact, begins to decline again in block 6.

The four remaining analyses summaries are presented in Tables 35 - 38. The means and standard errors are presented in Appendix G.

System Monitoring Task: Percent Correct Blocks 1-4

A 3-way ANOVA was performed on the dependent variable percent correct (%CORR). The three factors entered into the analysis were judgment type (JT) (absolute vs. comparative), block number (Blk) (1, 2, 3, 4) and condition (Cond)(ARMHH, ARMLH, CONHH, CONLH, CONHL). Again, the conditions ASMHH and ASMHL were not entered into the analysis, as no subject performance is reflected in these conditions during blocks 1 - 4.

The purpose of this analysis was to determine the possible effects of automating the resource (fuel) management task on performance within the system monitoring task. Additionally, the full manual control conditions were examined to determine the possible effects of high versus low levels of engagement within both the system monitoring and the resource (fuel) management tasks on performance within the system monitoring task.

Table 35: Summary Analysis for ARMHH System Monitoring False Alarms Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.67	14	3.97	0.168	0.688
Blk	5	0.47	70	1.37	0.342	0.886
JTxBlk	5	0.94	70	1.37	0.689	0.633

Table 36: Summary Analysis for ARMLH System Monitoring False Alarms Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	8.17	14	1.96	4.164	0.061
Blk	5	1.35	70	1.16	1.159	0.338
JTxBlk	5	0.74	70	1.16	0.637	0.672

Table 37: Summary Analysis for CONLH System Monitoring False Alarms Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	12.04	14	2.70	4.461	0.053
Blk	5	1.82	70	1.12	1.621	0.166
JTxBlk	5	0.89	70	1.12	0.796	0.556

Table 38: Summary Analysis for CONHL System Monitoring False Alarms Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1.04	14	0.39	2.612	0.128
Blk	5	0.19	70	0.52	0.371	0.867
JTxBlk	5	0.84	70	0.52	1.629	0.164

Table 39 displays the Summary Analysis for the System Monitoring Mean Percent Correct in blocks 1-4 across judgment types, block numbers, and conditions. As indicated in the summary table, there was a significant main effect for judgment type, $\underline{F}(1, 14) = 15.24$, $p < .01$, a significant main effect for block number, $\underline{F}(3, 42) = 9.97$, $p < .01$. There was also a significant two-way interaction between judgment type and block, $\underline{F}(3, 42) =2.96$, $p < .05$ and a significant two-way interaction between block number and condition, $\underline{F}(12, 168) = 2.36$, $p < .01$. Finally, the main effect for condition approached significance at $\underline{F}(4, 56) = 2.45$, $p = .057$.

The Greenhouse-Geisser procedure was applied to all significant effects. In the case of the main effect for block number, the degrees of freedom for error corrected from 3 to 2.45 and the degrees of freedom for error corrected form 42 to 34.27 The resulting p-value corrected form $p = .00004$ to $p = .0004$. In the case of the interaction between judgment type and block number, the degrees of freedom for effect corrected from 3 to 2.44 and the degrees of freedom for error corrected from 42 to 34.27. Although the p-value increased from $p = .043$ to $p = .065$, the importance of the interaction warranted its examination. Additionally, a less conservative adjustment, the Huynh-Feldt, did not change the degrees of freedom or p-value, so that the interaction remained significant. Finally, in the case of the two-way interaction between block number and condition, the degrees of freedom for effect corrected from 12 to 4.79 and the degrees of freedom for error corrected from 168 to 67.05. The p-value remained significant, correcting from $p = .008$ to $p = .049$.

Table 39: Summary Analysis for System Monitoring Mean Percent Correct Blocks 1-4

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	12.40	14	0.81	15.239	0.002
Blk	3	2.14	42	0.21	9.974	0.000
Cond	4	0.21	56	0.09	2.447	0.057
JTxBlk	3	0.64	42	0.21	2.962	0.043
JTxCond	4	0.07	56	0.09	0.826	0.514
BlkxCond	12	0.31	168	0.13	2.364	0.008
JTxBlkx Cond	12	0.07	168	0.13	0.508	0.907

Figure 35 shows the mean percent correct (means shown in arcsine correction) to be greater under the comparative versus absolute judgment type, as was predicted in Experiment I, Hypotheses # 3(a) (p. 32) and # 4 (p. 32).

Figure 36 shows a general decline in detection performance in the latter two blocks, reflecting a decline in vigilance, as was predicted. Post-hoc LSD tests revealed that blocks 1 and 2 did not differ significantly from each other, nor did blocks 3 and 4.

However, detection performance in block 1 was significantly better than in both blocks 3 ($p = .001$) and 4 ($p = .000$). Likewise, detection performance in block 2 was significantly better than in blocks 3 ($p = .000$) and 4 ($p = .000$).

Figure 37 shows better detection performance under conditions in which resource (fuel) management was automated. Post-hoc LSD tests revealed that detection performance was significantly better under ARMLH than its control, CONLH ($p = .009$), indicating that performance on the system monitoring task was aided by automation of the resource (fuel) management task, as predicted.

An examination of Figure 38, which displays the means and standard errors for the interaction between judgment type and block number, highlights two major findings. First, detection performance, overall, was better under the comparative judgment type, as indicated in the main effect for judgment type . This is supported in the post-hoc LSD results, which revealed that detection performance in block 1 of the absolute judgment type was significantly worse than in block 1 of the comparative

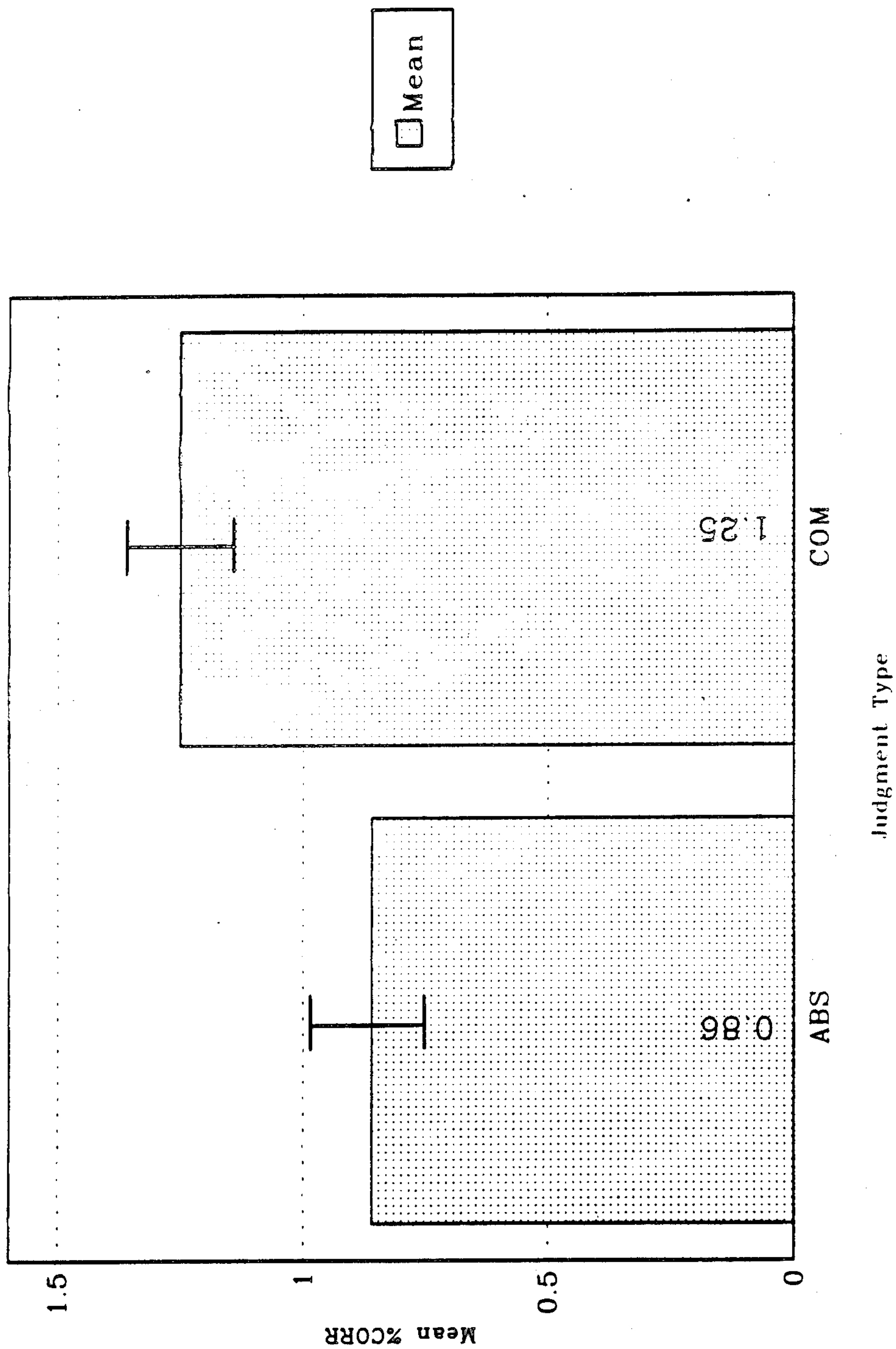


Figure 35. Judgment Type Means
System Monitoring Percent Correct Blocks 1-4
(Note: arcsine transformation values)

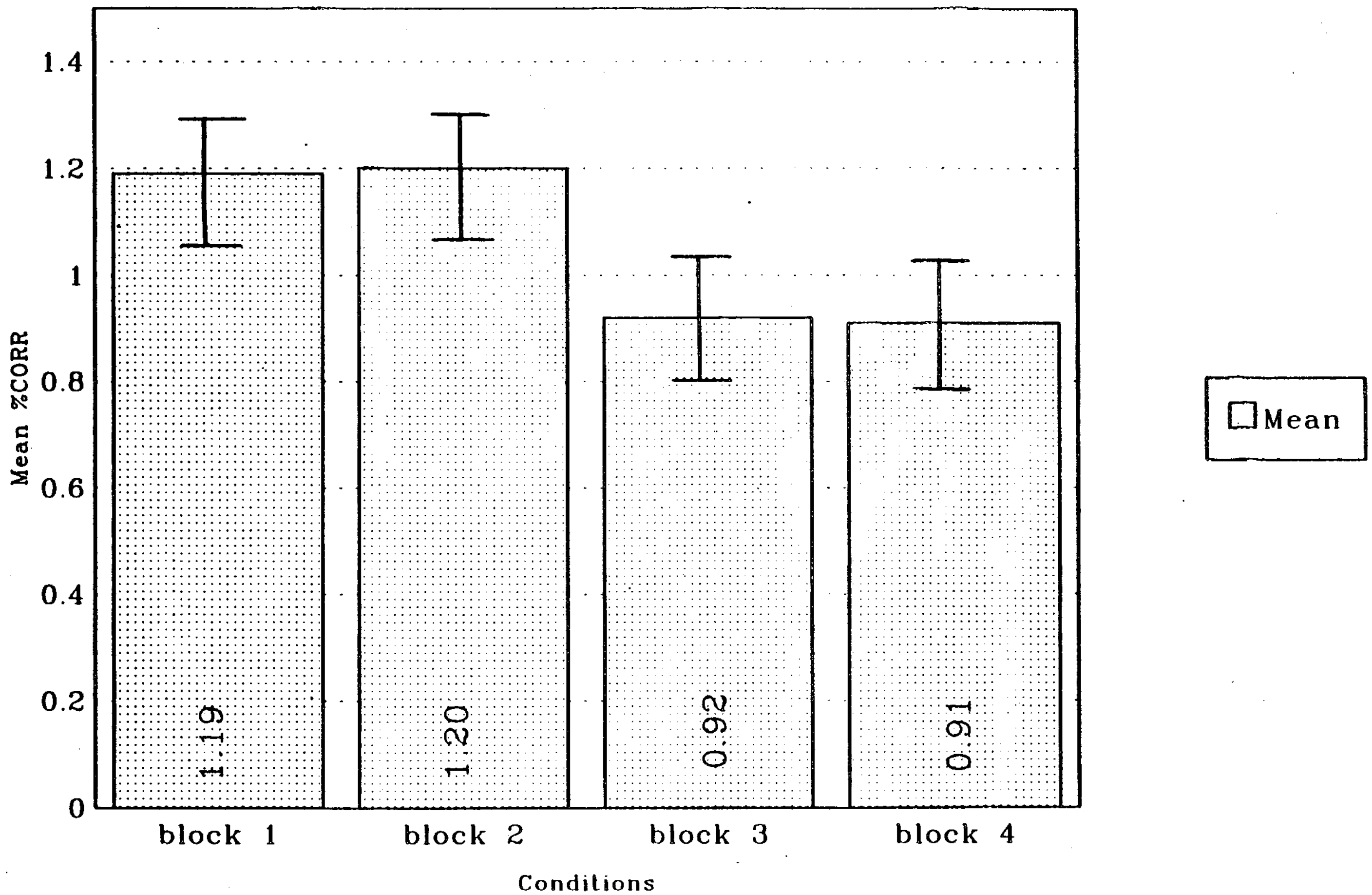


Figure 36: Block Means
System Monitoring Percent Correct Blocks 1-4
(Note: arcsine transformation values)

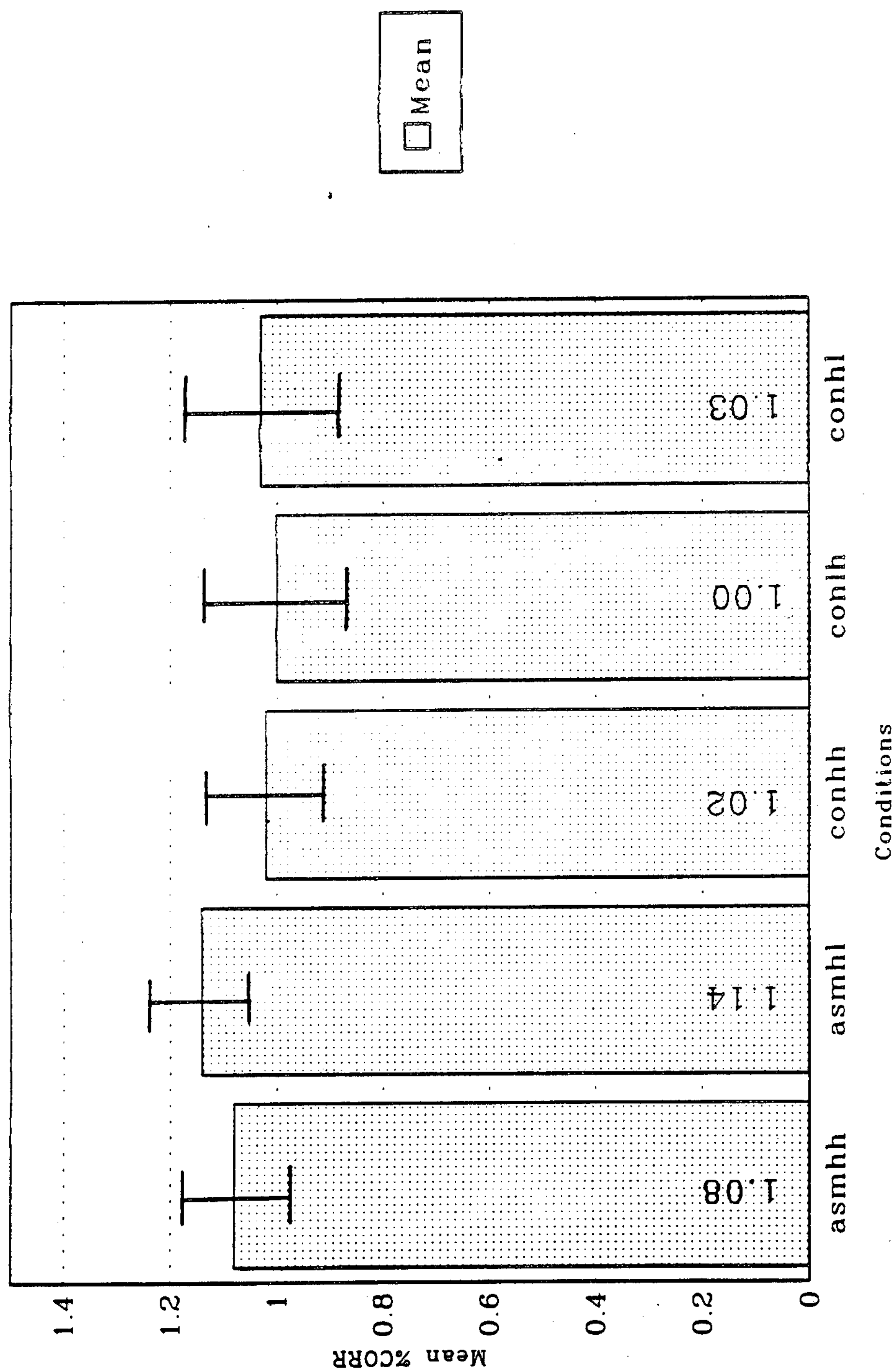


Figure 37. Condition Means
System Monitoring Percent Correct Blocks 1-4
(Note: arcsine transformation values)

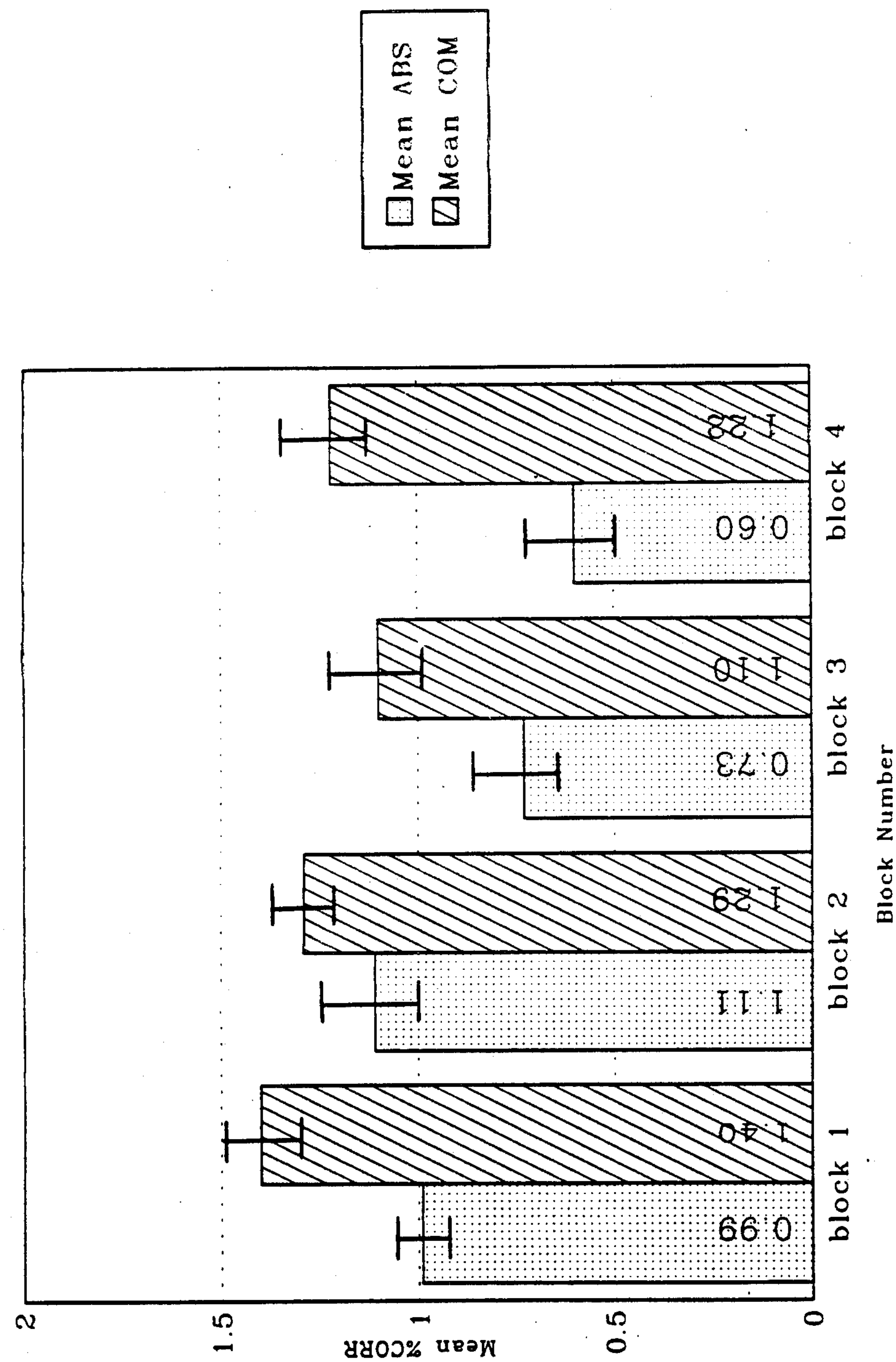


Figure 38: Judgment Type x Block Means
System Monitoring, Percent Correct Blocks 1-4
(Note: arcsine transformation values)

judgment type (p = .000). Additionally, detection performance in block 3 of the absolute judgment type was significantly worse than in block 3 of the comparative judgment type (p = .001). Finally, detection performance in block 4 of the absolute judgment type was significantly worse than in block 4 of the comparative judgment type (p = .000).

The second and more interesting finding revealed in Figure 38 is that the classic vigilance decrement occurs only under the absolute judgment type. This is further evidenced by the post-hoc LSD results, which revealed that detection performance in block 1 of the absolute judgment type was significantly better than in blocks 3 (p = .017) and 4 (p = .000). Likewise, detection performance in block 2 of the absolute judgment type was significantly better than in blocks 3 (p = .001) and 4 (p = .000). This was not quite the case, however, under the comparative judgment types, where detection performance was only significantly better in block 1 as compared to block 3 (p = .007). Interestingly, this indicates a greater vigilance decrement under the absolute judgment type. This would follow both Experiment I hypotheses # 3 (a) (p. 32) and # 4 (p. 32).

An examination of Figure 39, which displays the means and standard errors for the interaction between block number and condition, also highlights two major findings. The first is that detection performance is better overall in the initial blocks, reflecting the main effect for block. This is supported in the post-hoc LSD test, which revealed that, under the condition ARMHH, detection performance was significantly better in block 1 than in block 3 (p = .009) of the same condition. Detection performance was

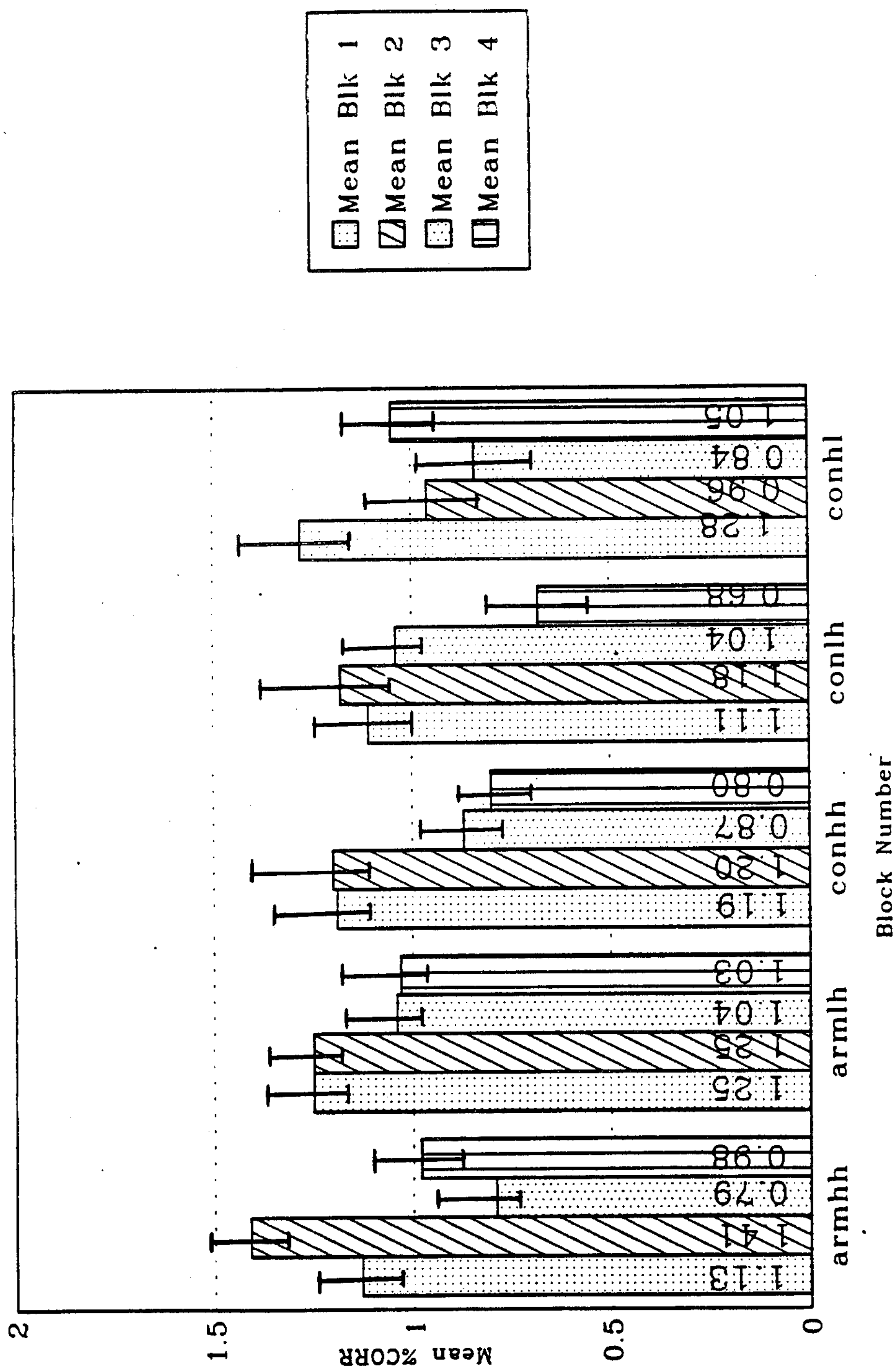


Figure 39: Block x Condition Interaction System Monitoring Percent Correct Blocks 1-4 (Note: arcsine transformation values)

also significantly better in block 2 than in blocks 3 (p = .000) and 4 (p = .001) of the same condition. Likewise, under CONHH, detection performance was significantly better in block 1 than in blocks 3 (p = .015) and 4 (p = .003) of the same condition. Detection performance was also significantly better in block 2 than in blocks 3 (p = .012) and 4 (p = .002) of the same condition. Detection performance was also significantly better in block 3 or CONHH than in block 4 of CONHH (p = .006). Furthermore, detection performance under CONLH was significantly better in block 1 than in block 4 (p = .001) of the same condition. Detection performance was also significantly better in block 2 than in block 4 (p = .000) of the same condition.

Finally, under CONHL, detection performance was significantly better in block 1 than in blocks 2 (p = .014) and 3 (p = .001).

The second finding is that the benefits to system monitoring detection are not really seen until block 3 and 4. The findings also indicate that the benefits to system monitoring seen in the latter blocks are due to two factors. The first, which is seen in block 3, is due to the engagement level of the resource (fuel) management task. Post-Hoc LSD tests revealed that detection performance in block 3 was significantly better under ARMLH than under ARMHH (p = .049). In other words, the engagement level of the resource (fuel) management task affected system monitoring detection performance, even though it was automated. Specifically, detection performance in the system monitoring task was significantly better when the engagement level in the automated resource (fuel) management task was low (6 pump failures per 10 minute block). Although this was not expected, it does not contradict any predictions. The second factor benefitting detection performance in the latter blocks of the system monitoring task is seen in block 4, where post-hoc LSD results indicate detection performance was significantly better under ARMLH as compared to its control, CONLH ($p = .007$). This supports the prediction that automating the resource (fuel) management task would enhance performance on the system monitoring task. It also highlights the unique interaction between block number and condition, in that the performance benefits due to automating resource (fuel) management are only seen in block 4.

System Monitoring Task: Percent Correct Block 5

A two-way ANOVA was performed on the dependent variable mean percent correct for block 5. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was twofold. First, to determine the possible effects of long-term automation and automation failure in the resource (fuel) management task on subsequent detection performance in the system monitoring task. Second, and more importantly, to determine the possible effects of long-term automation and automation failure in the system monitoring task on the subjects' performance in regaining manual control of that task.

Table 40 displays the summary analysis for the system monitoring task mean false alarm rate in block 5 across judgment types and conditions.

As indicated in the summary table, there was both a significant main effect for judgment type, $\underline{F} = 6.02$, $p < .05$ and a significant interaction between judgment

Table 40: Summary Analysis of System Monitoring Percent Correct Block 5

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3.90	14	0.65	6.017	0.028
Cond	6	0.32	84	0.14	2.322	0.040
JTxCon	6	0.03	84	0.14	0.220	0.969

type and condition, $\underline{F} = 2.32$, $p < .05$. The Greenhouse-Geisser was applied to the main effect for condition and adjusted the degrees of freedom for effect from 6 to 3.92 and the degrees of freedom for error from 84 to 54.82. The p-level adjusted from p = .040 to p = .068.

Regarding the significant main effect for judgment type, an examination of the means in Figure 40 shows that, as predicted, the detection rate was significantly better under the comparative judgment type.

With respect to the main effect for condition, the means displayed in Figure 41 show that subjects performed worse (lower percent correct detection rate) immediately following automation of the system monitoring task. Post-hoc LSD analyses indicated that detection performance in block 5 of the condition ASMHH was significantly worse than in block 5 of ARMHH (p = .013) and block 5 of its control, CONHH (p = .049). Furthermore, detection performance in block 5 of ASMHL was significantly worse than in block 5 of its control, CONHL (p = .049). These results indicate that regaining manual control of the system monitoring task was more difficult following automation of system monitoring than if that task had not been automated.

System Monitoring Task: Percent Correct Block 6

A two-way ANOVA was performed on the dependent variable mean percent correct for block 6. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and condition (Cond)(ARMHH, ARMLH, ASMHH, ASMHL, CONHH, CONLH, CONHL). The purpose of this analysis was to determine any lingering effects of automation and automation failure of the resource (fuel)

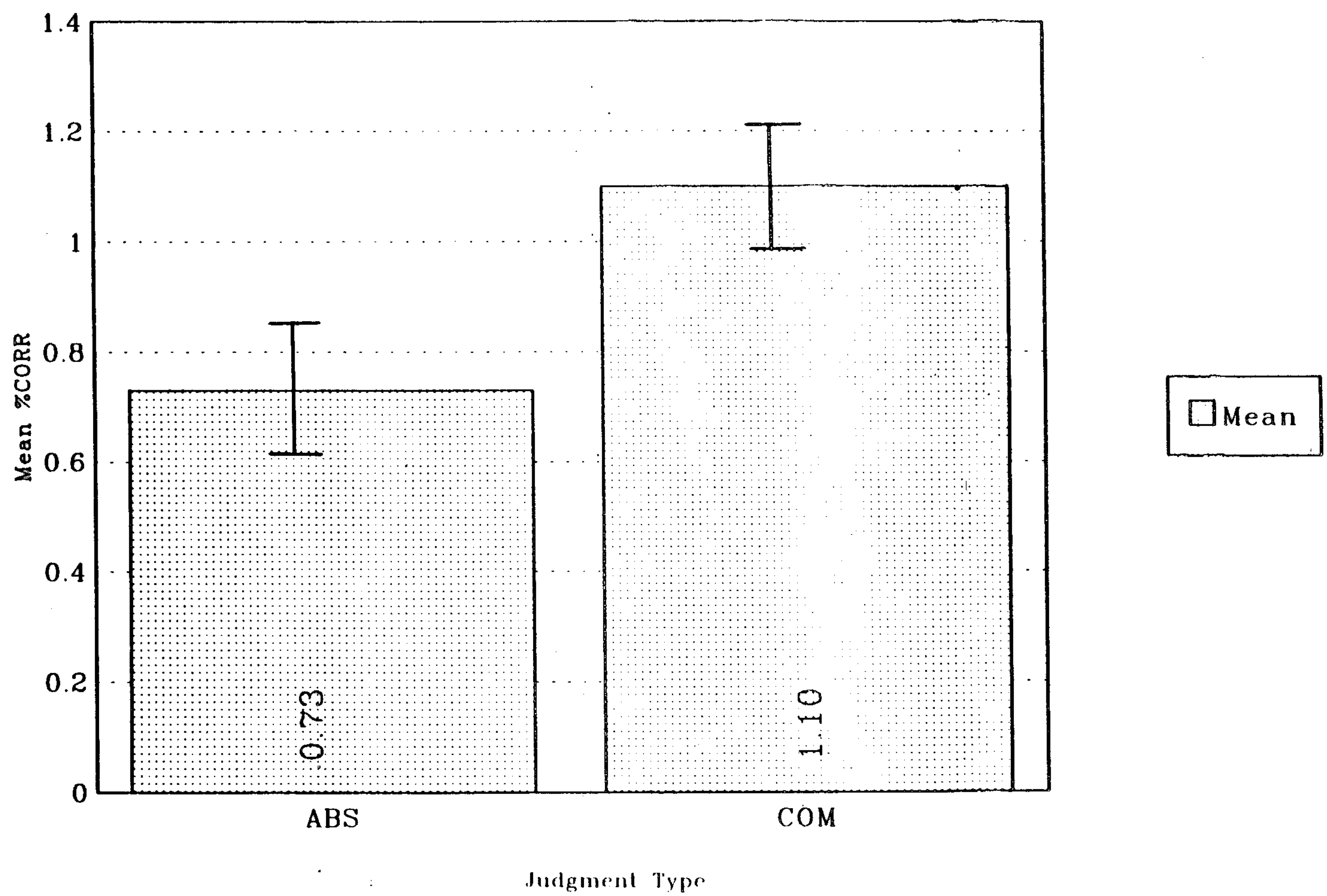


Figure 40: Judgment Type Means
System Monitoring Percent Correct Block 5
(Note: arcsine transformation values)

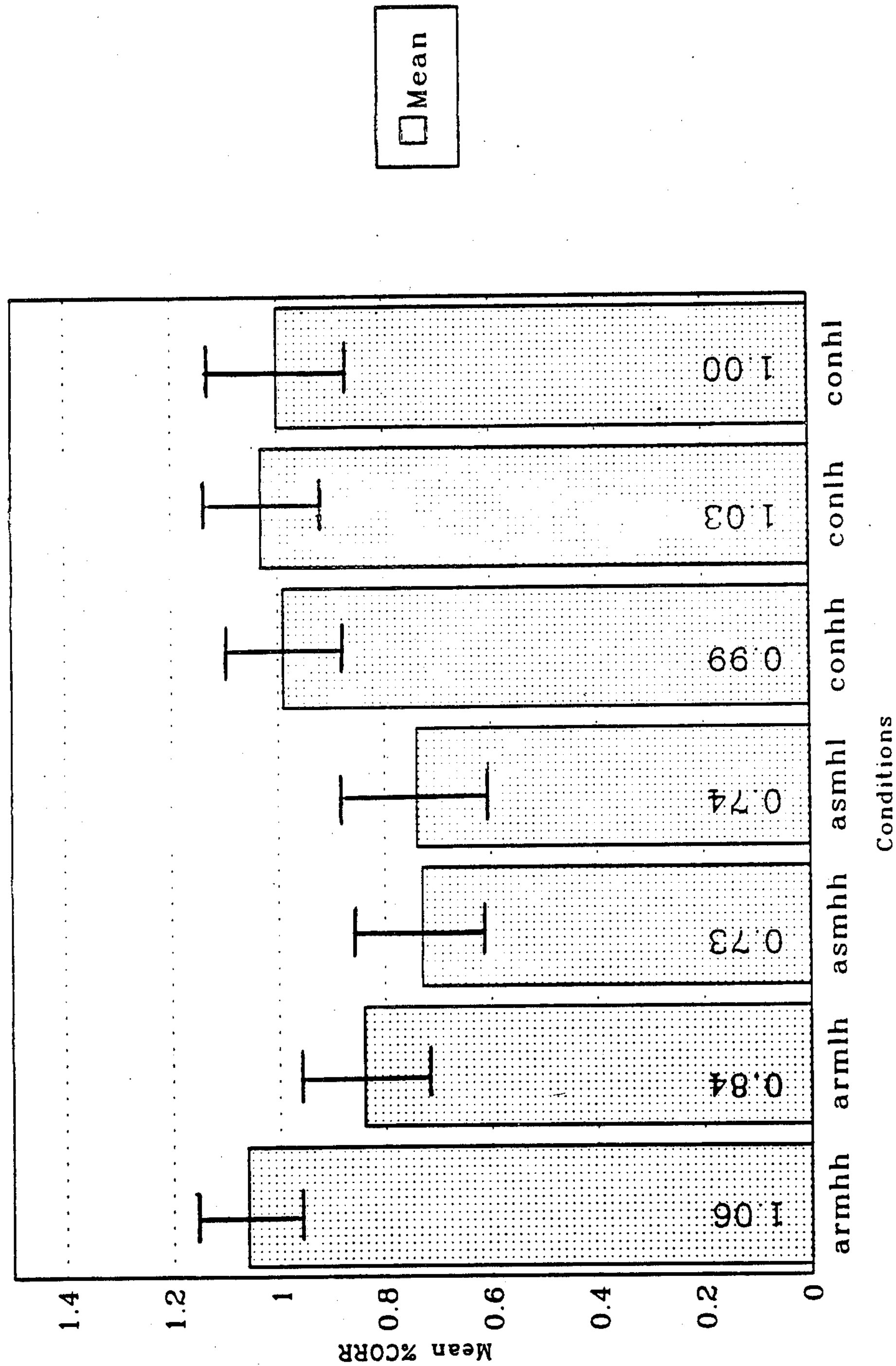


Figure 41: Condition Means
System Monitoring Percent Correct Block 5
(Note: arcsine transformation values)

management task, and especially the system monitoring task, on detection rate in the system monitoring task.

Table 41 displays the summary analysis for the system monitoring task mean percent correct in block 6 across judgment types and conditions.

As indicated in the summary table, there was a significant main effect for judgment type, $\underline{F} = 5.79$, $p < .05$ and a significant main effect for condition, $\underline{F} = 3.22$, $p < .01$. The Greenhouse-Geisser adjustment was applied to the main effect for condition, resulting in adjustment of degrees of freedom for effect from 6 to 3.74 and adjustment of degrees of freedom for error from 84 to 52.43. The p-level adjusted from $p = .007$ to $p = .019$. The means for the two judgment types, displayed in Figure 42, show that, as predicted in Experiment I, Hypotheses # 3(a) (p. 32) and # 4 (p. 32), the detection rate was significantly better under the comparative judgment type.

Regarding the significant main effect for the condition, an examination of the means in Figure 42a indicates the beneficial effects on system monitoring detection performance related to automating the resource (fuel) management task continued into block 6. Post-hoc LSD tests showed that detection performance in block 6 under ARMLH was significantly better than in block 6 of its control, CONLH ($p = .013$). However, the detrimental effects on detection performance related to automating the system monitoring task also continued into block 6. Post-hoc LSD tests indicated that detection performance in block 6 under ASMHL was significantly worse than in block 6 of its control, CONHL ($p = .009$). Furthermore, an interesting, unpredicted effect

Table 41: Summary Analysis for System Monitoring Percent Correct Block 6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3.43	14	0.59	5.788	0.031
Cond	6	0.48	84	0.15	3.219	0.007
JTxCon	6	0.18	84	0.15	1.179	0.326

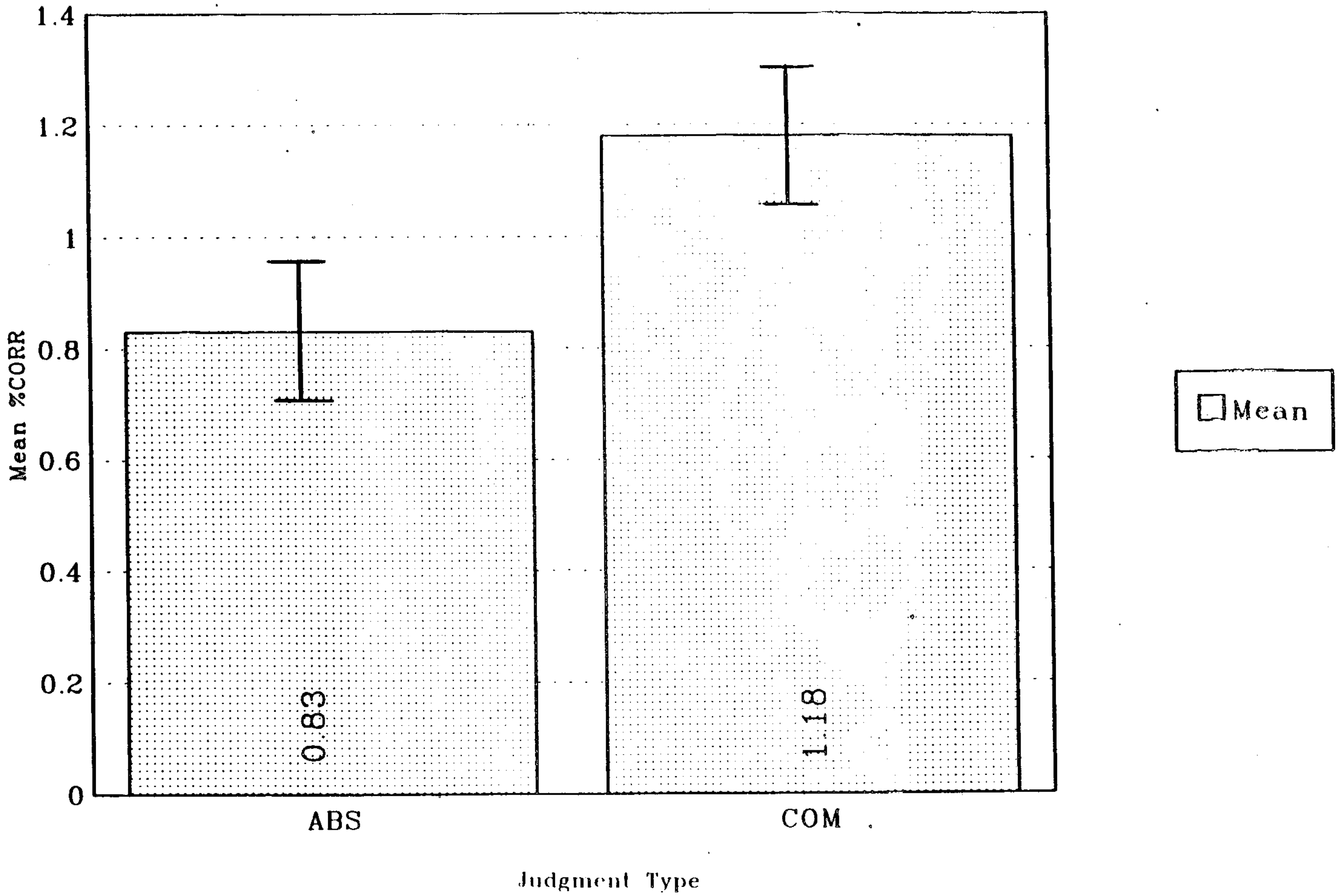


Figure 42: Judgment Type Means
System Monitoring Percent Correct Block 6
(Note: arcsine transformation values)

612A

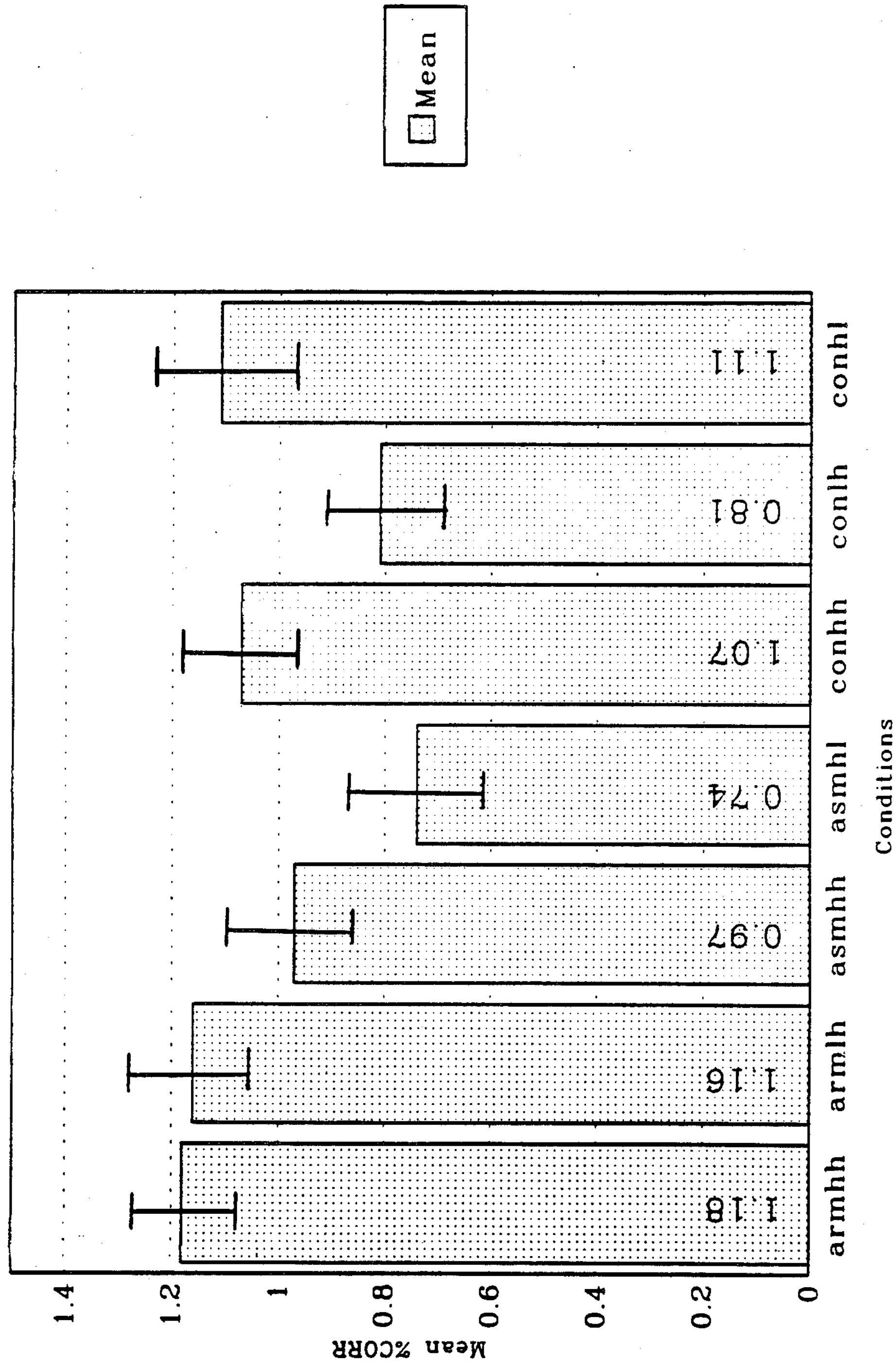


Figure 42a: Condition Means
System Monitoring Percent Correct Block 6
(Note: arcsine transformation values)

from engagement level was indicated in the post-hoc LSD. Detection performance in block 6 of CONLH (engagement level of the resource [fuel] management task low, engagement level of the system monitoring task high) was significantly worse ($p = .031$) than in block 6 of CONHL (engagement level of resource [fuel] management high, engagement level of system monitoring task low) and significantly worse ($p = .056$) than in block 6 of CONHH (engagement level of both tasks high). This effect was not predicted and it did not appear in block 5. If it is genuine, it may simply indicate that with the lower engagement level in the resource (fuel) management task, vigilance went down towards the end of the session. This would support predictions of greater demand from resource (fuel) management task.

System Monitoring Task: Percent Correct Blocks 1-6

In order to examine any general trends in detection rate over an entire condition session, a separate 2-way ANOVA was performed on block 1-6 mean percent correct for each of the conditions, except for the two in which system monitoring was automated during blocks 1-4 (ASMHH and ASMHL). These were not entered because no human system monitoring performance data was gathered during blocks 1-4 of these conditions, as this task was automated at that time. The two factors entered into the analysis were judgment type (JT) (absolute vs. comparative) and block number (Blk) (1, 2, 3, 4, 5, 6).

Of the conditions examined, three had significant main effects for block. The specific results of the separate analyses follow.

Considering the condition ARMHH, Table 42 displays the summary analysis for the percentage of correct detections in blocks 1-6. As indicated in the summary table, there was no significant main effect for block number. This is depicted visually in Figure 43, which shows no particular fluctuation pattern in the block means. There was, however, a significant main effect for judgment type, $\underline{F}$ (1, 14) = 17.44, $p < .01$. The judgment type means in Figure 44 show that, as expected, subjects scored a higher percentage of correct detections under the comparative judgment type.

Considering the condition ARMLH, Table 43 displays the summary analysis for the percentage of correct detections in blocks 1-6. As indicated in the summary table, there was a significant main effect for block number, $\underline{F}$(5, 70) = 5.32, $p < .01$. The significant main effect for block was subjected to the Greenhouse-Geisser adjustment. This adjusted the degrees of freedom for effect from 5 to 3.26 and the degrees of freedom for error from 70 to 45.62. The effect remained significant, with the p-level adjusting from p = .0003 to p = .003.

An examination of the block means in Figure 45 shows fairly uniform performance across blocks 1 - 4, followed by a sudden drop in performance in block 5. Post-hoc LSD tests revealed that block 5 was significantly worse than blocks 1 (p = .000), 2 (p = .000), 3 (p = .000), 4 (p = .003) and 6 (p = .001). This seems to indicate that the automation failure of the resource (fuel) management task, which occurred at the onset of block 5 in ARMLH, had a detrimental effect on detection performance within the system monitoring task. Although this outcome was not specifically predicted, it implies that failure of automation in the resource (fuel) management task is detrimental

Table 42: Summary Analysis for ARMHH System Monitoring Percent Correct Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3.83	14	0.22	17.439	0.001
Blk	5	0.07	70	0.11	0.670	0.648
JTxBlk	5	0.06	70	0.11	0.561	0.730

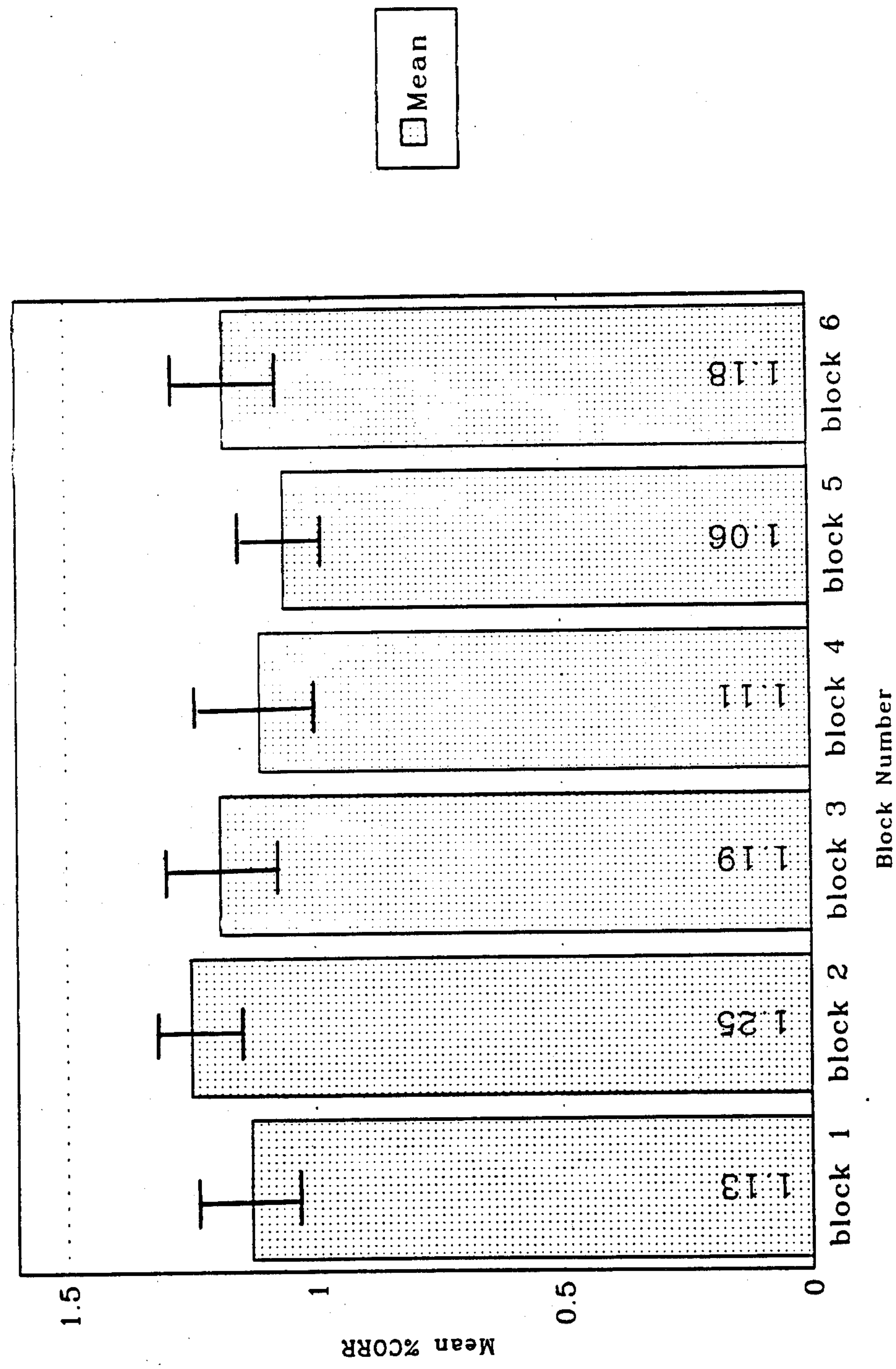


Figure 43. ARMHH Block Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

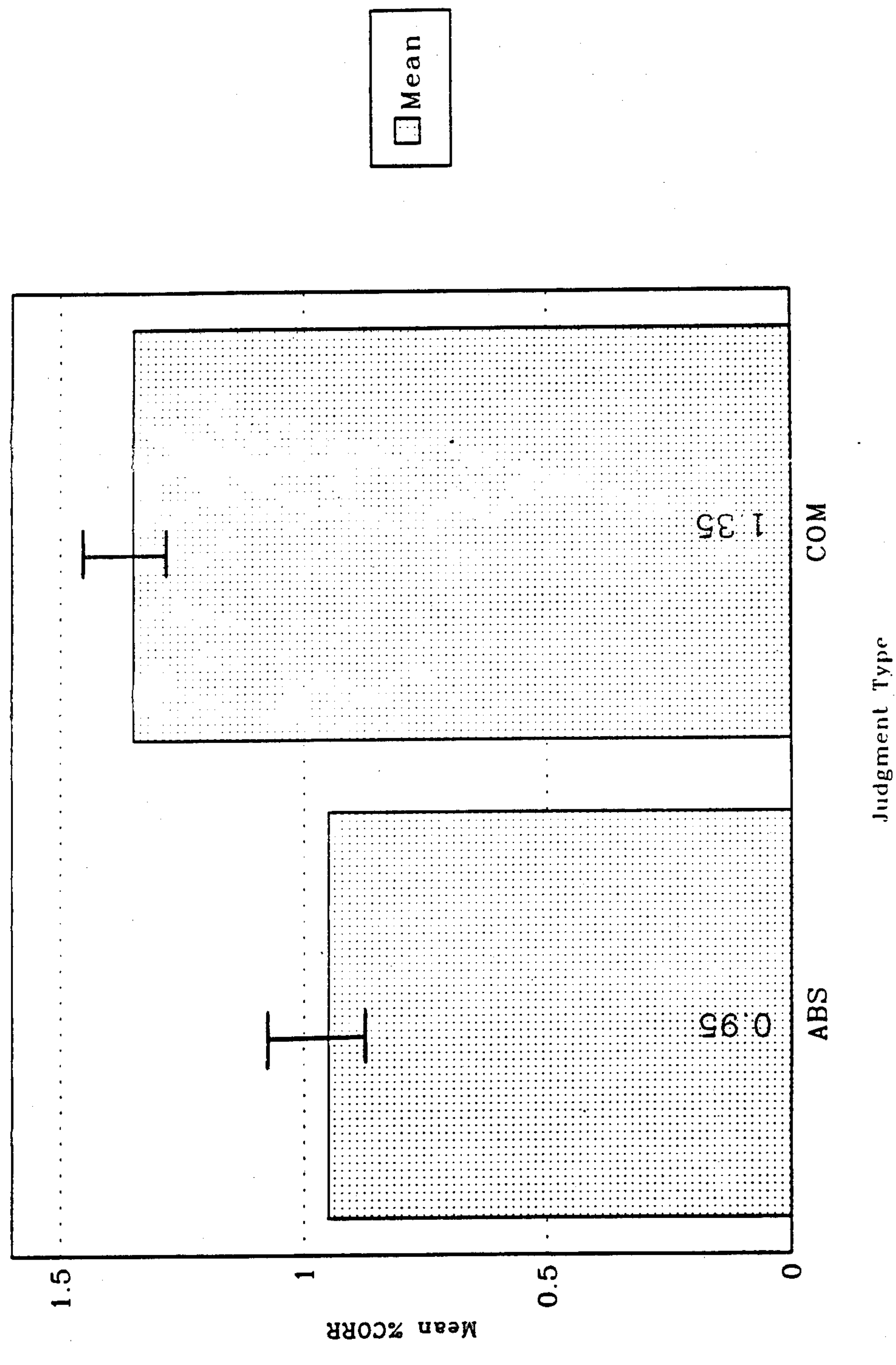


Figure 44: ARMHH Judgment Type Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

Table 43: Summary Analysis of ARMLH System Monitoring Percent Correct Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.80	14	0.40	2.011	0.178
Blk	5	0.60	70	0.11	5.318	0.000
JTxBlk	5	0.05	70	0.11	0.466	0.800

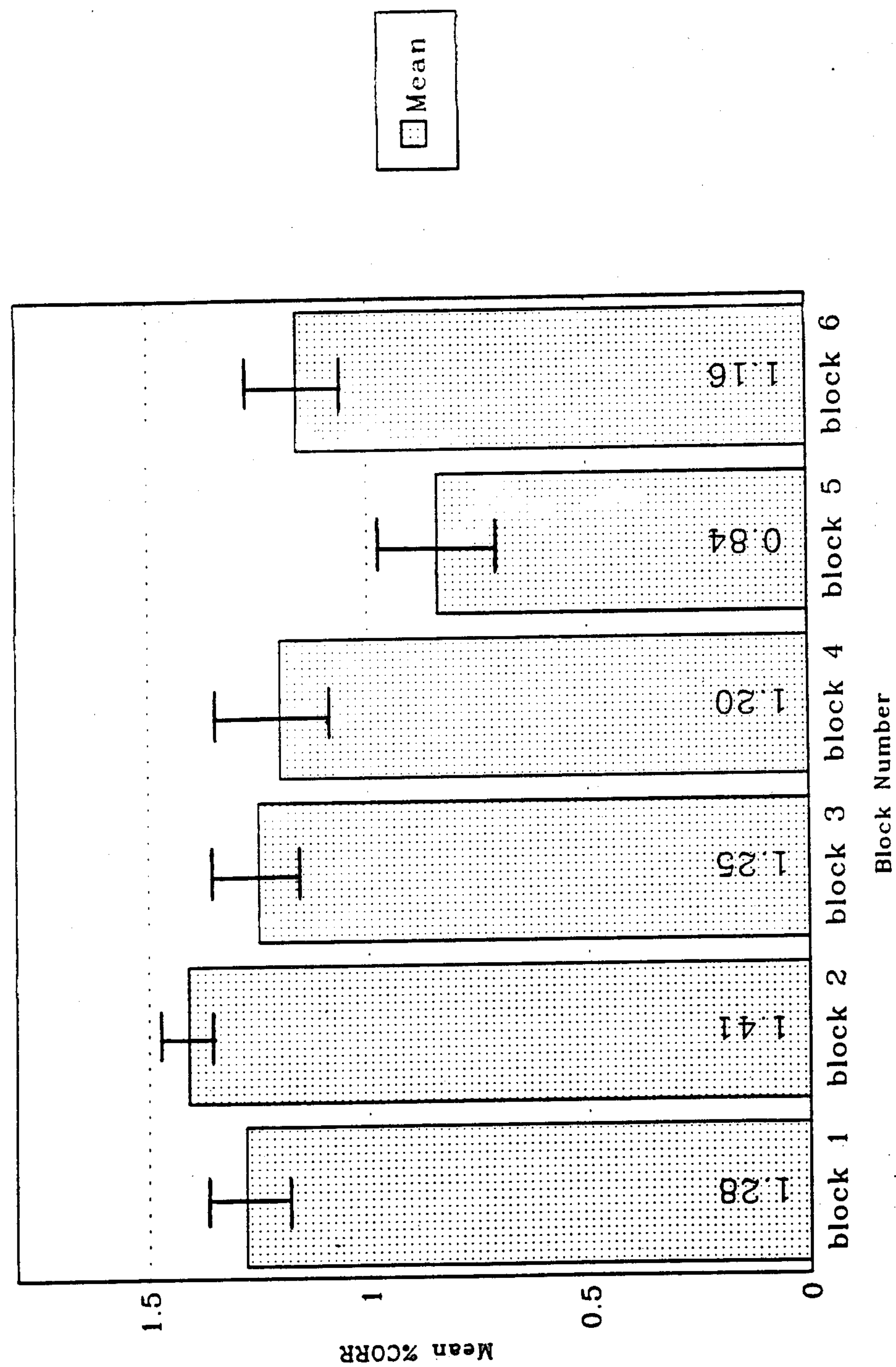


Figure 45: ARMLH Block Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

to overall task performance. When considering that subject performance on the resource (fuel) management task also dropped significantly following automation failure of the resource (fuel) management task, the data implies that the subject's overall task performance following automation of resource (fuel) management is seriously detrimentally affected.

Considering the condition CONHH, Table 44 displays the summary analysis for the percentage of correct detections in blocks 1-6. As indicated in the summary table, there was a significant main effect for judgment type, $\underline{F}(1, 14) = 7.13$, $p < .05$. Figure 46 shows that, as expected, the mean for the comparative judgment type is significantly higher than the mean for the absolute judgment type. Table 44 also shows a significant main effect for block number, $\underline{F}(5, 70) = 2.61$, $p < .05$. The Greenhouse-Geisser procedure was applied to the main effect for block number. It adjusted the degrees of freedom for effect from 5 to 3.56 and the degrees of freedom for error from 70 to 49.91. The effect remained significant, with the p-level adjusting from $p = .032$ to $p = .046$. An examination of the block means in Figure 47 shows highest performance in block 1, followed by a drop in performance through block 3, and an increase in performance in block 4, which was relatively stable through block 6. Post-Hoc LSD tests indicated that block 1 was significantly better than block 3 ($p = .001$), while block 3 was significantly worse than blocks 4 ($p = .031$) and 6 ($p = .016$). This fluctuation in the full manual control condition appears to be a variation of the classic vigilance phenomena, with subjects' performance showing the typical decline over a period of about 15 to 20 minutes, but followed with a "wake-up" period

Table 44: Summary Analysis for CONHH System Monitoring Percent Correct Blks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	3.45	14	0.48	7.127	0.018
Blk	5	0.28	70	0.11	2.613	0.032
JTxBlk	5	0.07	70	0.11	0.672	0.646

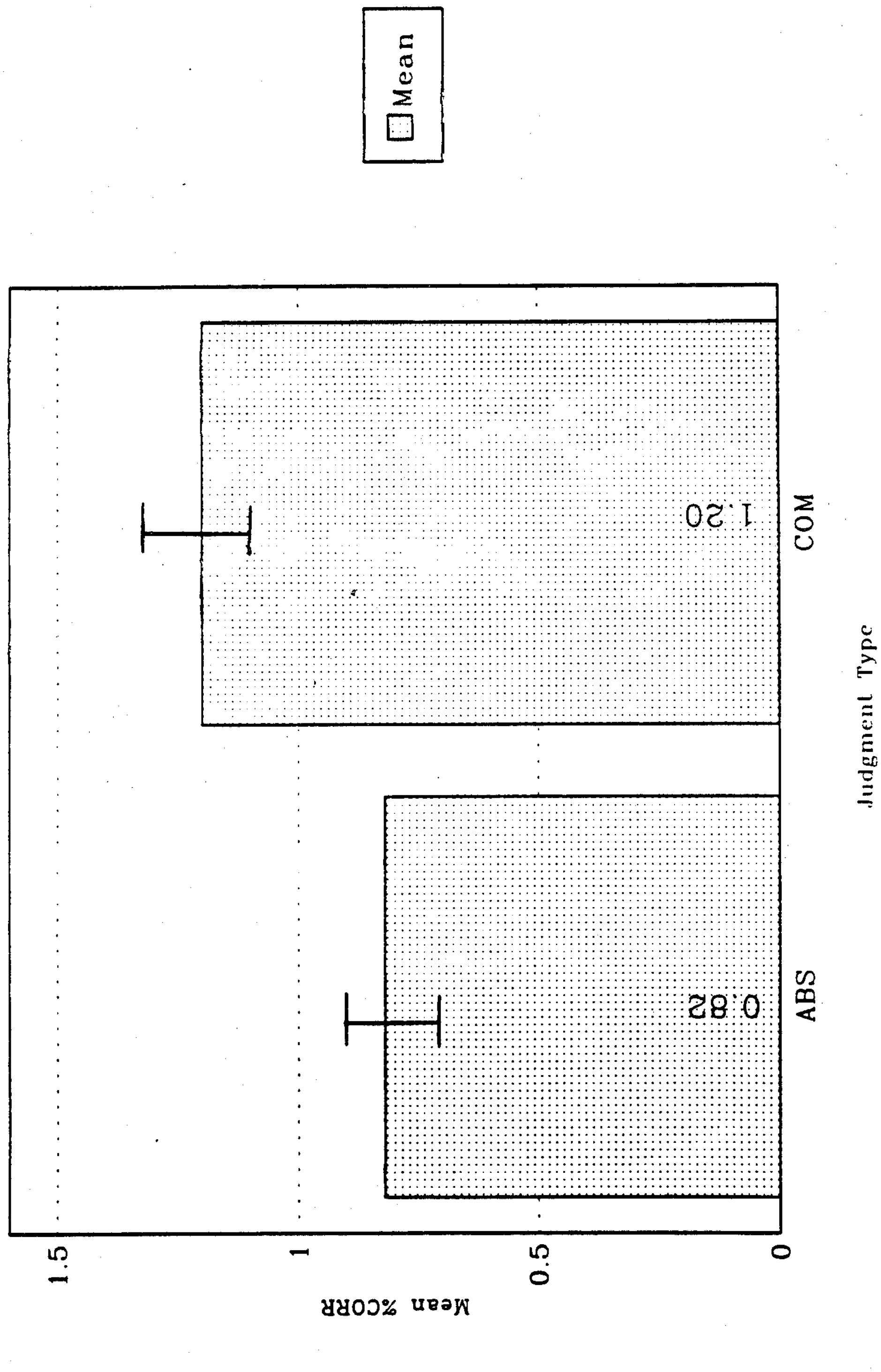


Figure 46: CONHH Judgment Type Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

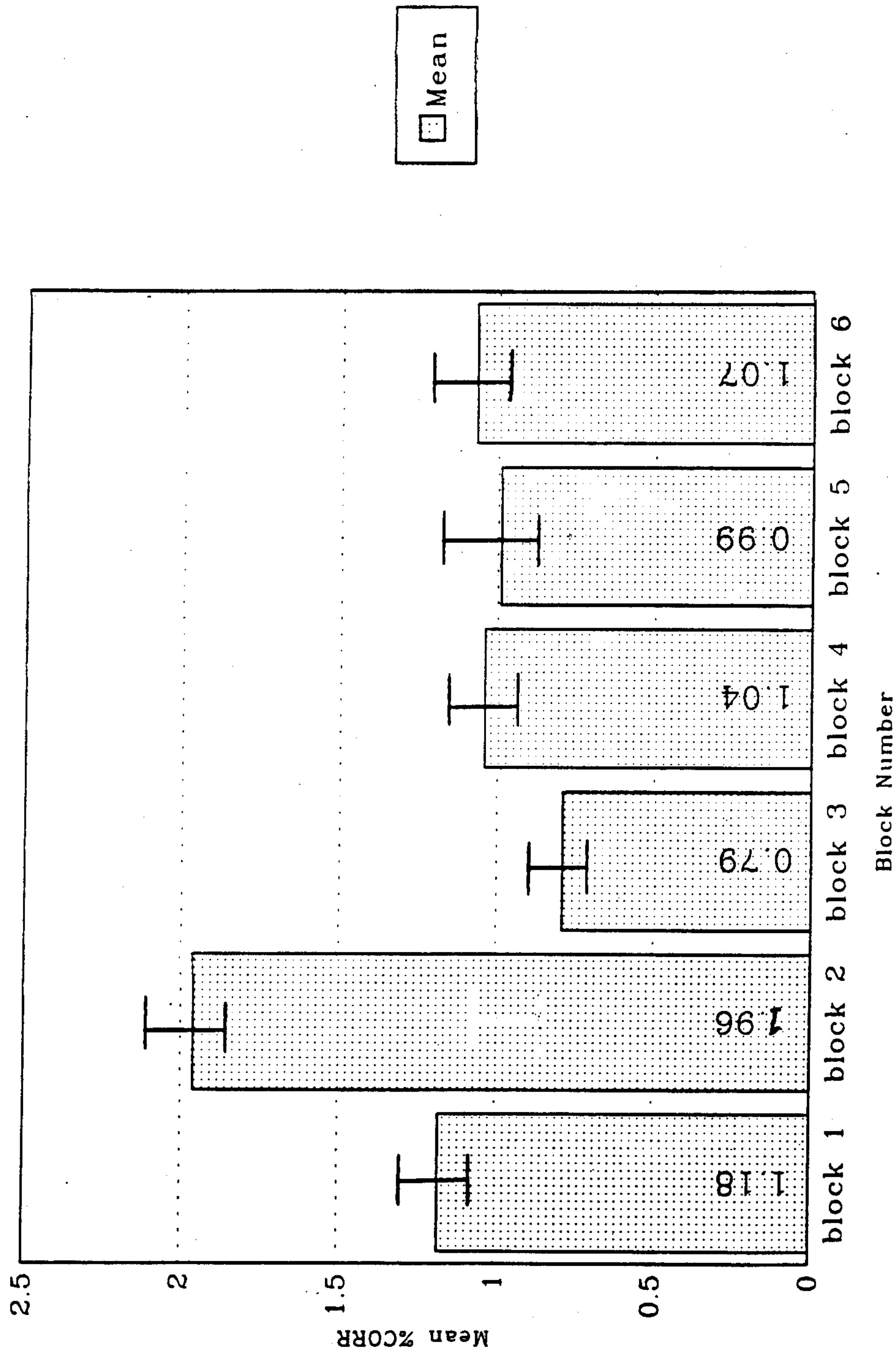


Figure 47: CONIIH Block Means
System Monitoring Percent Correct Blocks 1 6
(Note: arcsine transformation values)

of about 10 to 15 minutes. It is also not unlike the fluctuation in performance in the resource (fuel) management task during blocks 1-6 of the full manual controls.

Considering the condition CONLH, Table 45 displays the summary analysis for the percentage of correct detections in blocks 1-6. As indicated in the summary table, there were no significant main effects or interactions in this analysis. Figure 48 shows the means for judgment type and indicates that performance was better (though not significantly) under the comparative judgment type. Furthermore, Figure 49 shows the means for block, and indicates a fluctuating pattern which is somewhat different from the other two controls. Again, this was neither significant, nor predicted.

Considering the condition CONHL, Table 46 displays the summary analysis for the percentage of correct detections in blocks 1-6. As indicated in the summary table, there was a significant main effect for judgment type, $\underline{F}(1, 14) = 11.99$, $p < .01$. Figure 50 shows the pervasive effect that the mean percent correct is higher under the comparative judgment type.

There was also a significant main effect for block number, $\underline{F}(5, 70) = 2.66$, $p < .05$. The Greenhouse-Geisser procedure was applied to the main effect for block number. It adjusted the degrees of freedom for effect from 5 to 3.06 and the degrees of freedom for error from 70 to 42.82. The effect remained marginally significant, with the p-level adjusting from $p = .032$ to $p = .060$. An examination of the block means in Figure 51 shows a pattern almost identical to that seen under CONHH. The highest performance is in block 1, followed by a drop in performance through block 3,

Table 45: Summary Analysis for CONLH System Monitoring Percent Correct Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	2.88	14	0.80	3.604	0.078
Blk	5	0.16	70	0.10	1.613	0.168
JTxBlk	5	0.16	70	0.10	0.531	0.191

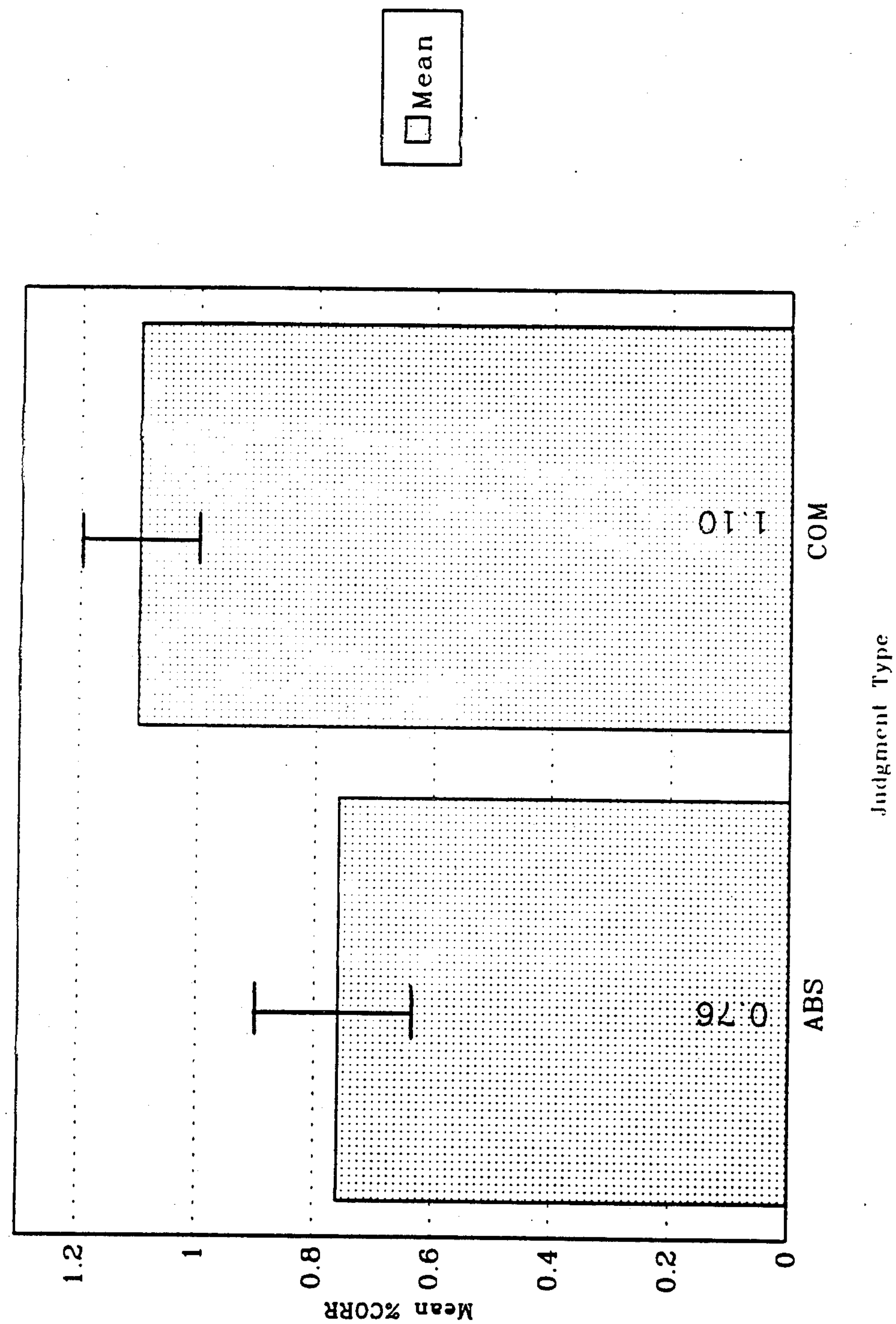


Figure 48: CONLH Judgment Type Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

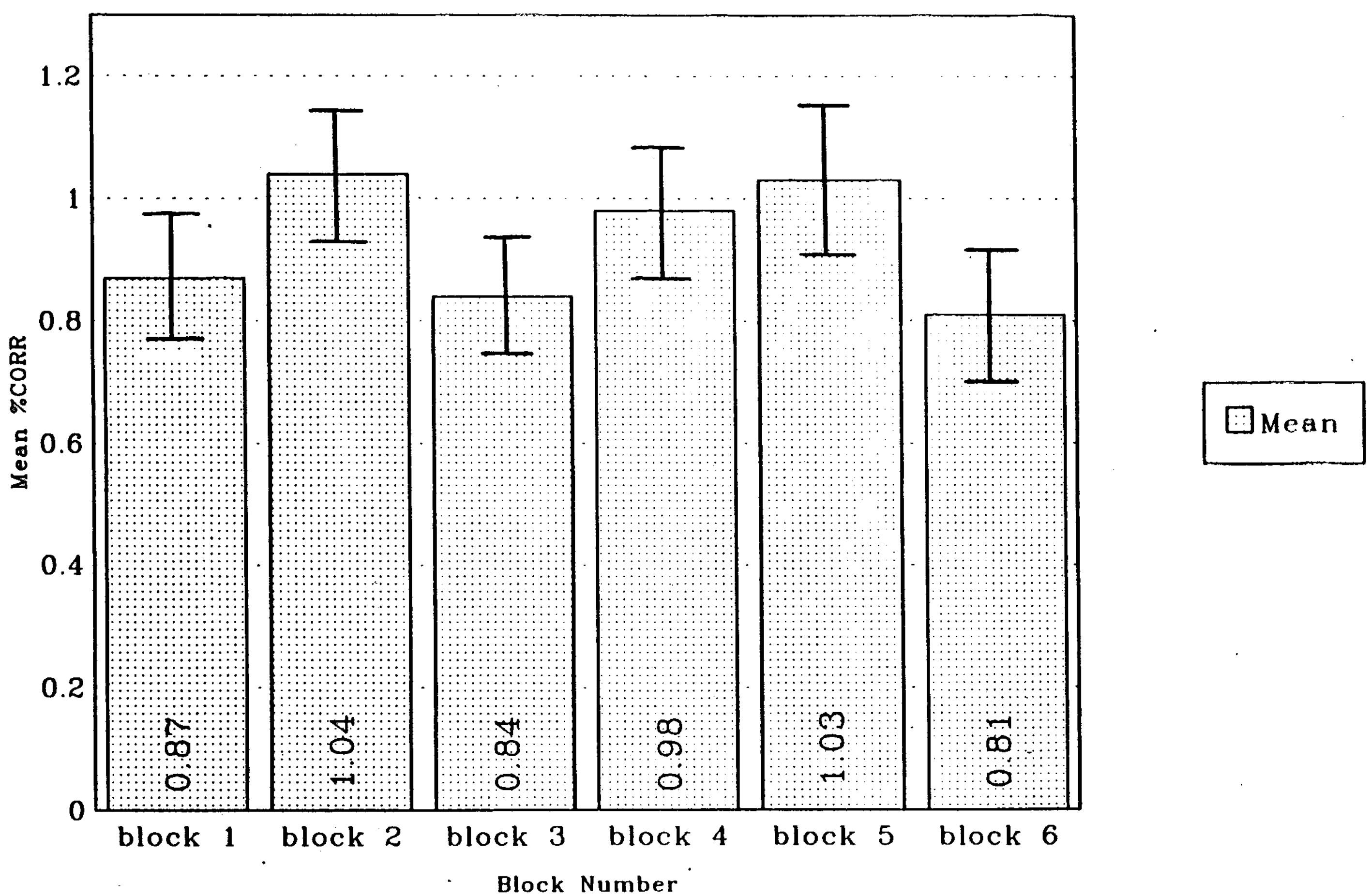


Figure 49 CONLH Block Means
System Monitoring Percent Correct Blocks 1 - 6
(Note: arcsine transformation values)

Table 46: Summary Analysis for CONHL System Monitoring Percent Correct Blocks 1-6

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	6.85	14	0.57	11.986	0.004
Blk	5	0.45	70	0.17	2.662	0.029
JTxBlk	5	0.14	70	0.17	0.852	0.518

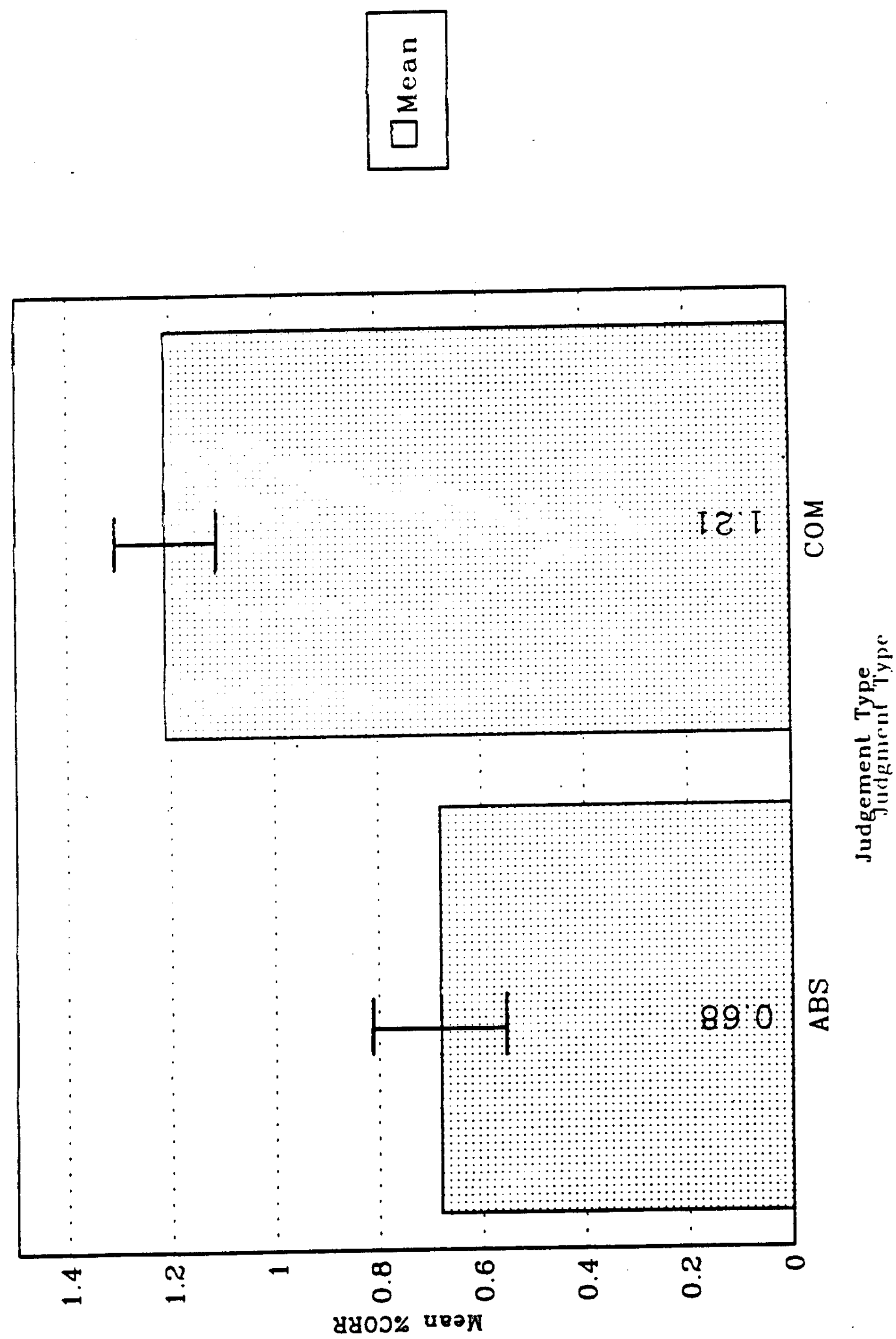


Figure 50: CONHIL Judgment Type Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

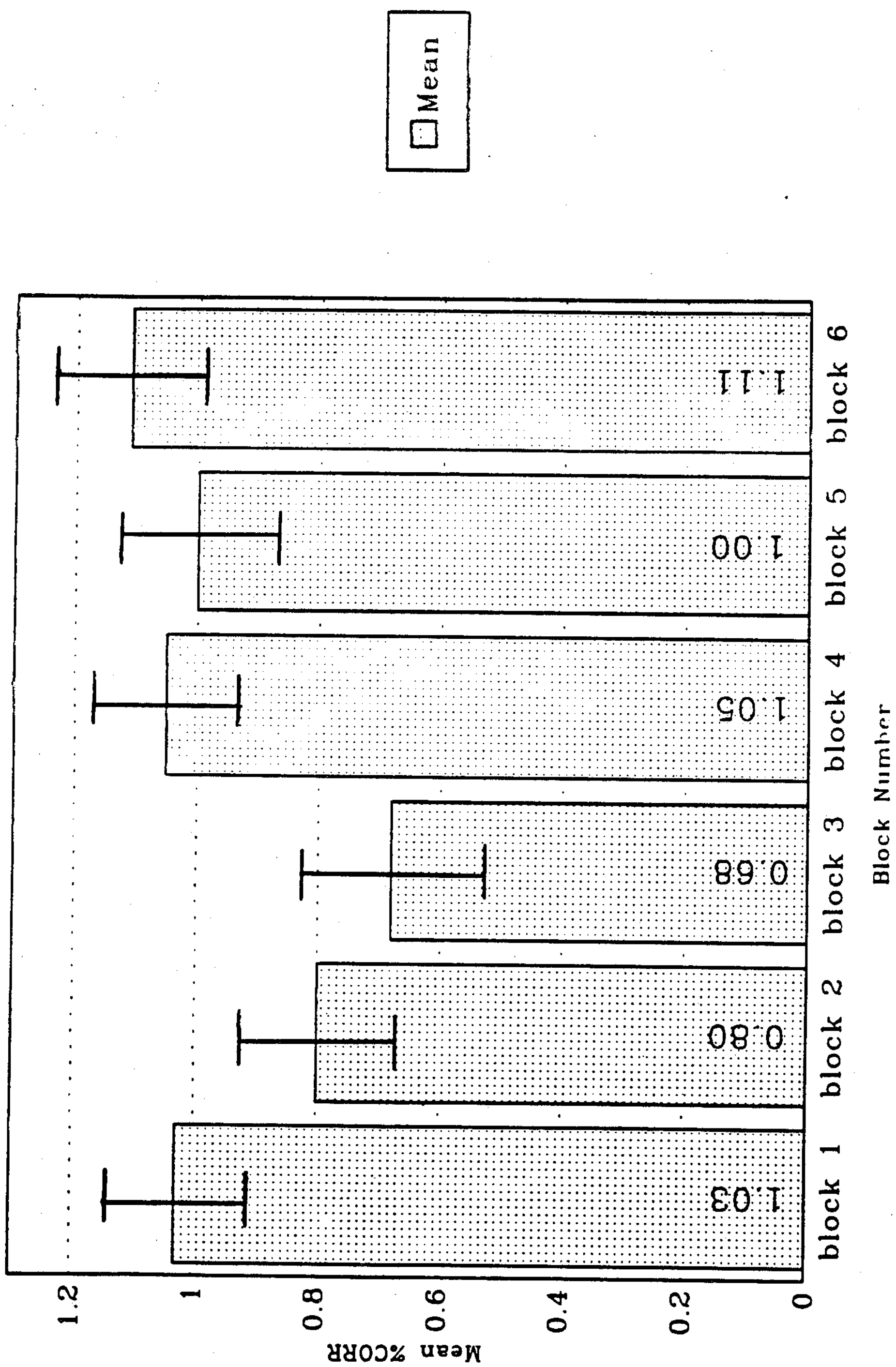


Figure 51. CONHIL Block Means
System Monitoring Percent Correct Blocks 1-6
(Note: arcsine transformation values)

and an increase in performance in block 4, which was relatively stable through block 6. Post-Hoc LSD tests indicated that block 1 was significantly better than block 3 ($p = .017$), while block 3 was significantly worse than blocks 4 ($p = .014$), 5 ($p = .031$) and 6 ($p = .004$). Again, this fluctuation in the full manual control condition appears to be somewhat related to classic vigilance phenomena. It is also not unlike the fluctuation in performance in the resource (fuel) management task during blocks 1-6 of the full manual control conditions.

Secondary Analysis: Workload

All NASA-TLX scores for each subject were entered into the Complete Statistical Software package (CSS), version 3.1 (Statsoft, 1992), and a series of ANOVAs conducted. Analyses focused upon main effects and interactions of judgment types (absolute vs. comparative) and conditions, as with the performance data.

There were no significant effects for the overall score in the NASA-TLX. The summary analysis is presented in Table 47, with the means for judgment type and condition shown in Figures 52 and 53, respectively.

There was a significant main effect for condition on the frustration subscale. The summary analysis for this effect is displayed in Table 48, and the condition means are presented graphically in Figure 54.

As indicated in the graph, the frustration level was highest for the two control conditions in which one of the tasks had a low engagement level. This difference was significant for ASMHL.

Table 47: Summary Analysis for TLX Overall Score

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1522.938	14	1134.131	1.343	0.266
Cond	6	126.577	84	103.643	1.221	0.304
JTxCond	6	47.750	84	103.643	0.461	0.835

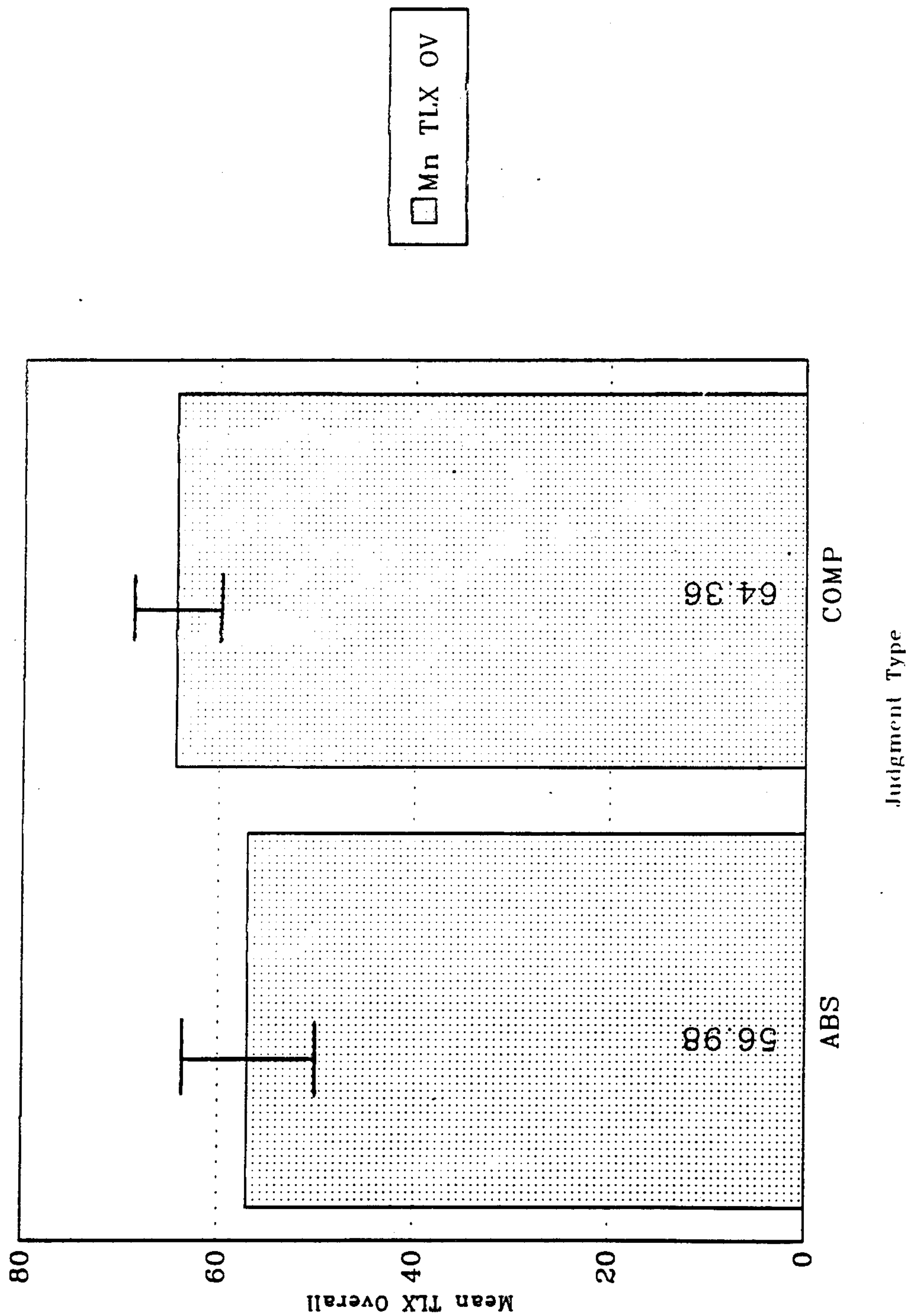


Figure 52: Judgment Means - TLX Overall Score

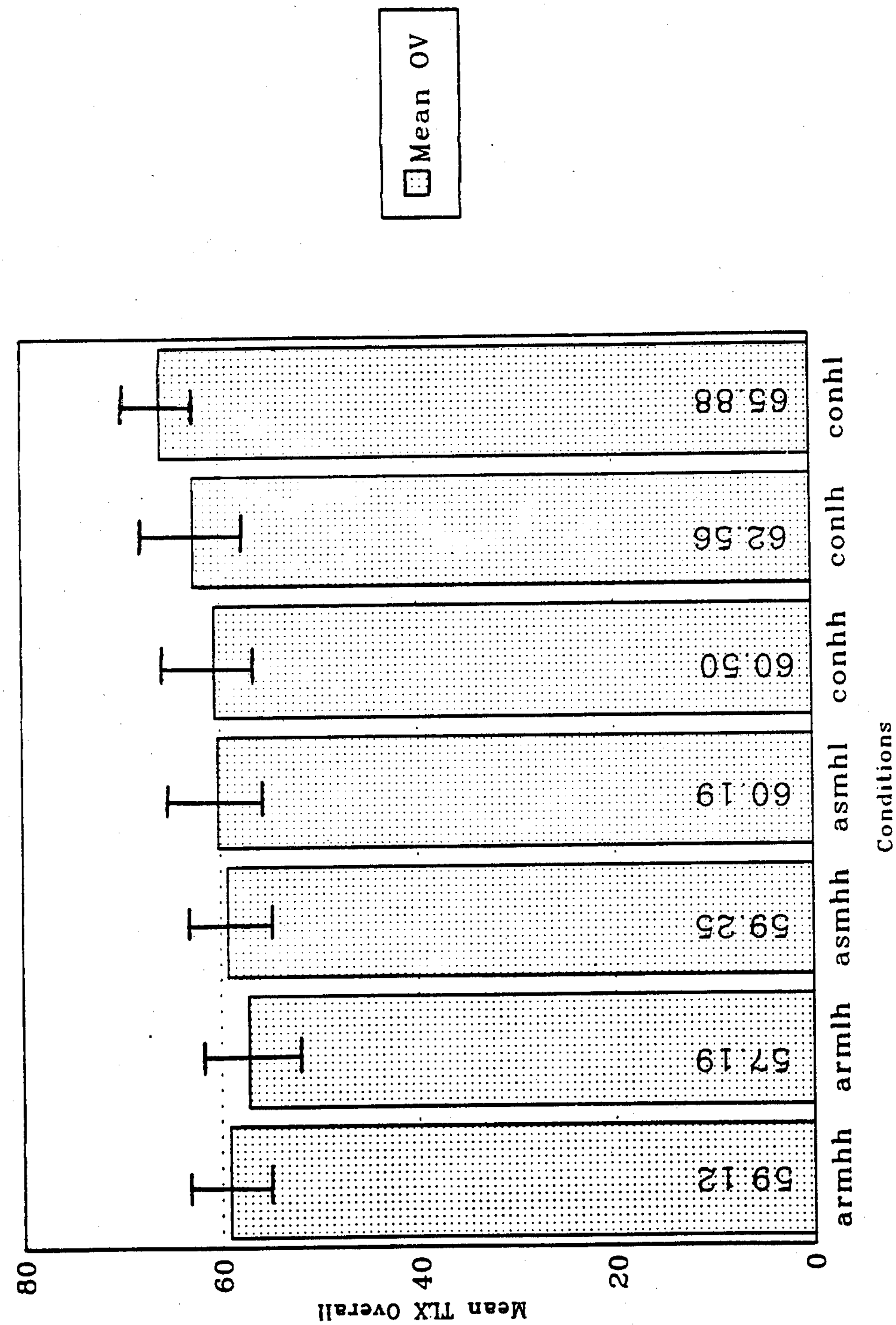


Figure 53. Condition Means
TLX Overall Score

Table 48: Summary Analysis for TLX Frustration Subscale

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	43.948	14	76.830	0.572	0.462
Cond	6	23.795	84	9.276	2.565	0.025
JTxCond	6	5.143	84	9.276	0.554	0.765

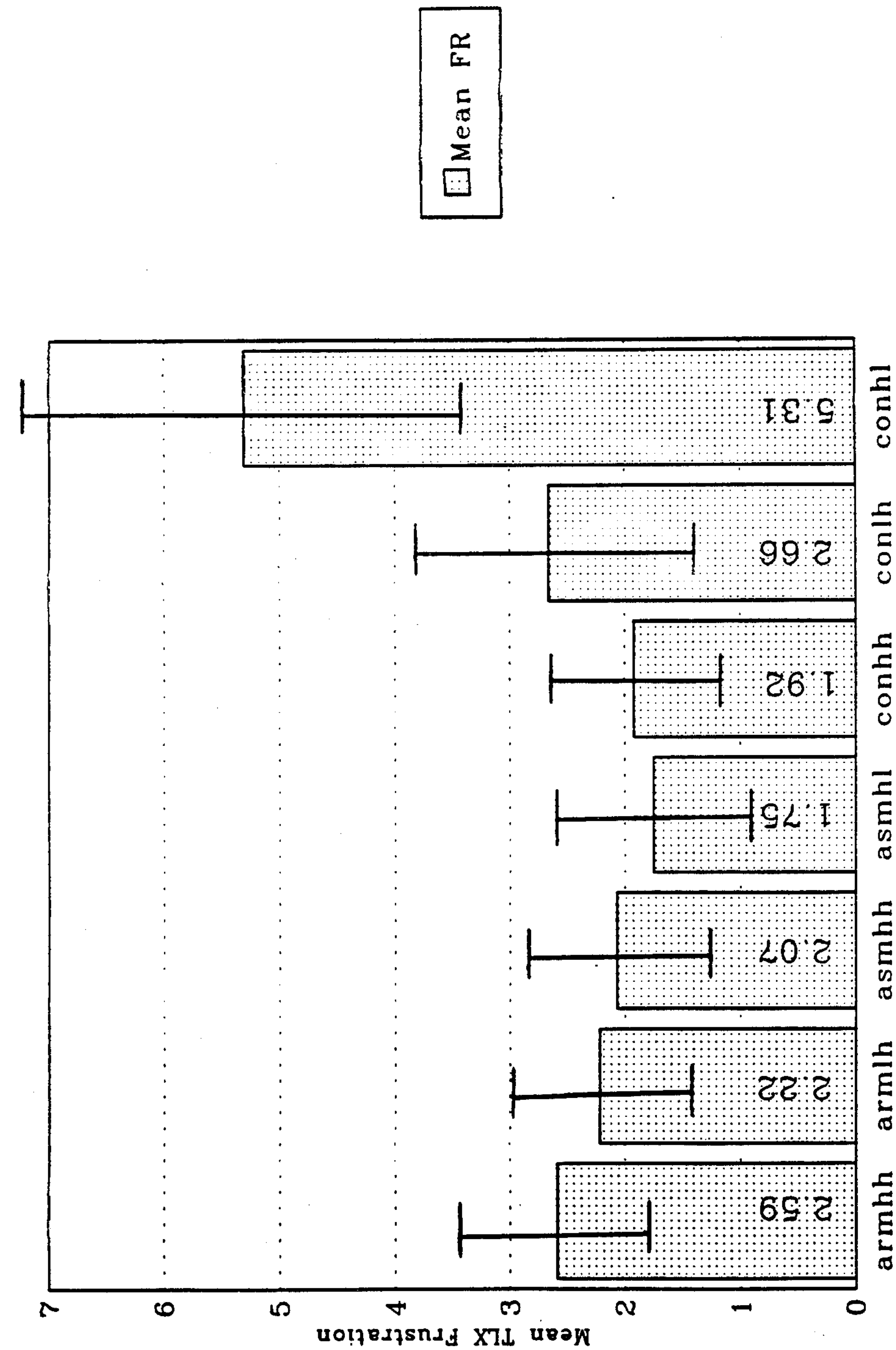


Figure 54 Condition Means - TLX Frustration Score

Study II Analysis: S.A. Data

Again, for each of the 3 conditions in the second study (automated resource (fuel) management, automated system monitoring, and full manual control), each subject was run twice. During each trial, the subject was asked 2 questions pertaining to the perception of events (Level I SA) in the resource (fuel) management task and in the system monitoring task, 4 questions pertaining to the meaning of information (Level II SA) in the resource (fuel) management task and the system monitoring task, and 1 question pertaining to the future of events (Level III SA) in the resource (fuel) management task and the system monitoring task (only one Level III question was asked because the system monitoring task did not lend itself to such a question type). Each of these questions was then scored on the basis of percent correct (subject answers compared with computerized account of actual occurrences). In this manner, a percent correct score was attained for each question type within each task of interest (resource [fuel] management and system monitoring) for both trials in each condition for each subject.

All the scores in the two trial-per-condition sets were then averaged to attain a mean percent correct for each question type for each of the tasks of interest for each condition for each subject (one subject in the absolute judgment type group attrited during the second full manual control trial, and this score was therefore based on only one trial).

All means were then subjected to an arcsine transformation and entered into CSS 3.1 (Statsoft, 1992). A series of one-way ANOVAs compared like question types of automated groups to their full manual control groups, as well as to the opposing automated group (i.e., automated resource [fuel] management [ARM] vs. automated system monitoring [ASM]).

Level I (Perception) SA Questions about Resource (Fuel) Management

Table 49 displays the summary analysis for Level I questions about the resource (fuel) management task. The two conditions entered in this analysis were automated resource (fuel) management (ARM) and automated system monitoring (ASM). As indicated in Table 49, there were no significant main effects or interactions for Level I(Perception) SA questions about the resource (fuel) management task. However, the main effect for judgment type approached significance, $\underline{F}(1, 14) = 4.38$, $p = .055$. Contrary to what was predicted in Experiment II, Hypothesis # 1 (pg. 34), Figure 55 indicates that subjects had a higher level of perception SA for events in the resource (fuel) management task under the absolute judgment type. On the other hand, although the main effect for condition was not significant, Figure 56 indicates that slightly more questions pertaining to a perception of events in the resource (fuel) management task were answered correctly while the system monitoring task was automated, as opposed to while the resource (fuel) management task was automated.

Table 50 again displays a summary analysis for Level I (Perception) questions about the resource (fuel) management task. In this analysis, however, the two conditions entered were automated resource (fuel) management (ARM) and the full

Table 49: Summary Analysis for Level I (Perception) Questions about Resource (Fuel) Management: ARM versus ASM

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.87	14	0.20	4.382	0.055
Cond	1	0.41	14	0.33	1.249	0.282
JTxCond	1	0.07	14	0.33	0.223	0.644

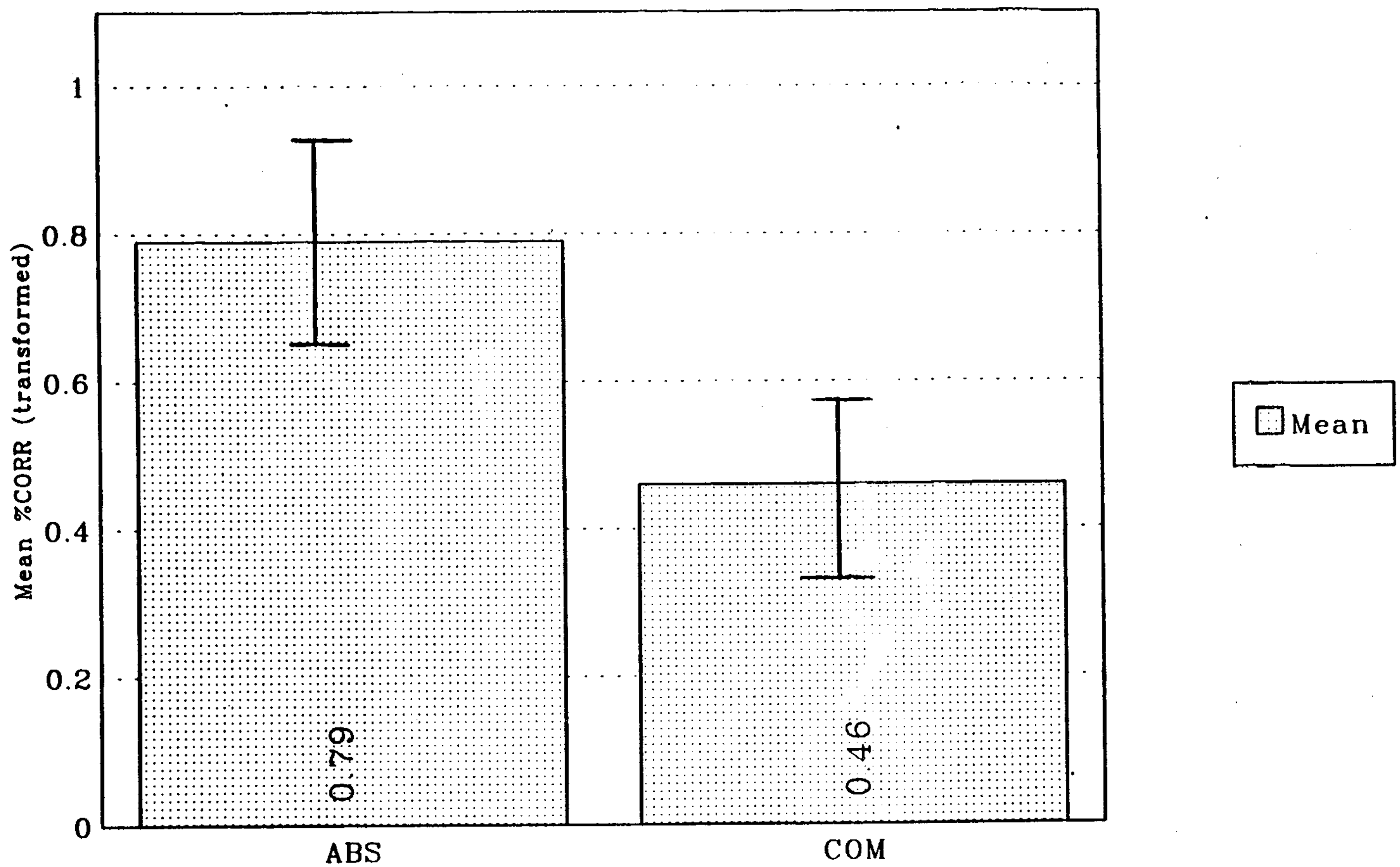


Figure 55: Judgment Type Means
Level I (Perception) SA of Resource (Fuel) Management
ARM vs ASM

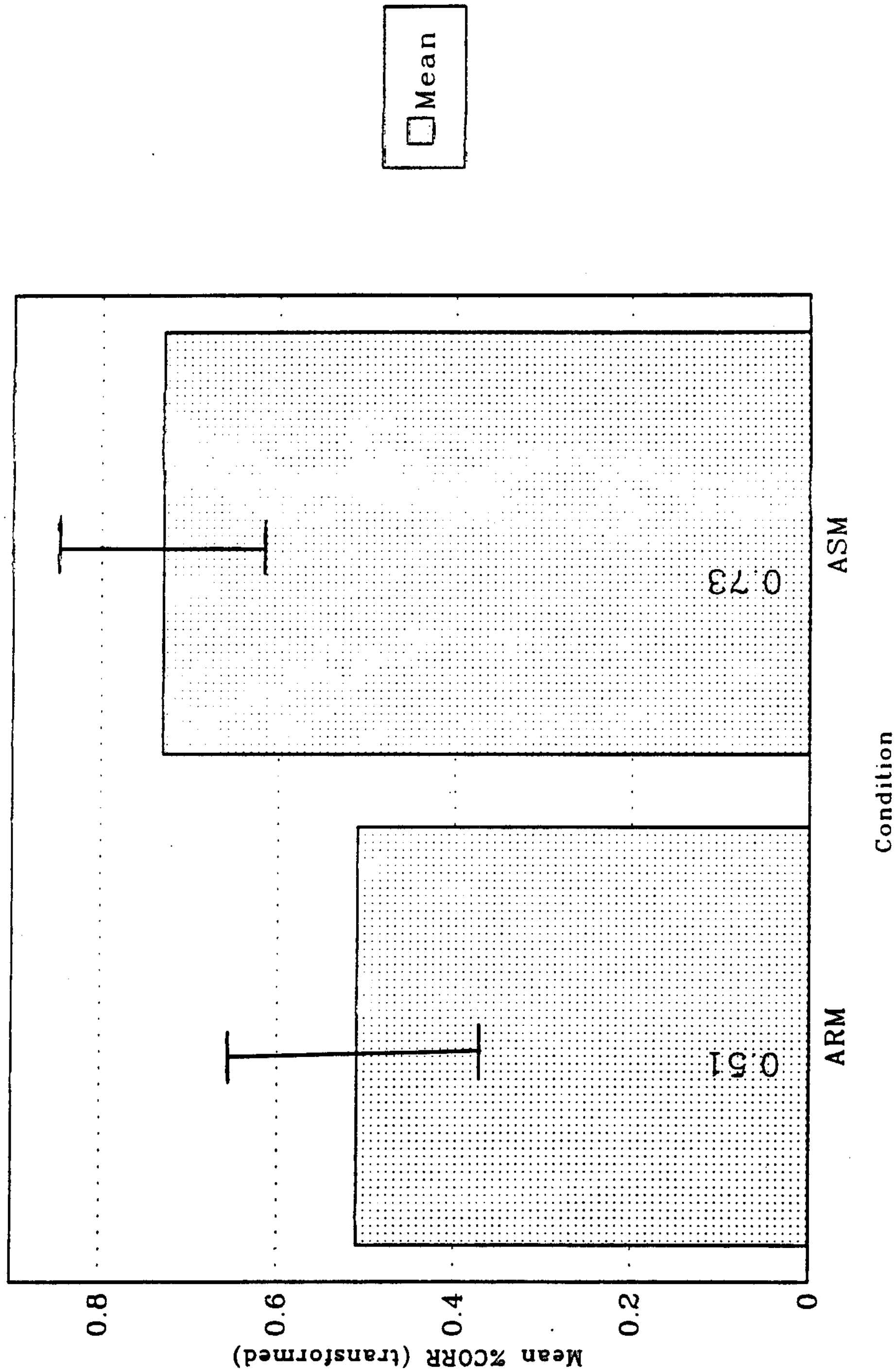


Figure 56. Condition Means
Level 1 (Perception) SA of Resource (Fuel) Management
ARM vs ASM

Table 50: Summary Analysis for Level I (Perception) Questions about Resource (Fuel) Management: ARM versus CON

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.00	14	0.36	0.001	0.971
Cond	1	0.14	14	0.19	0.746	0.402
JTxCond	1	0.47	14	0.19	2.523	0.134

manual control condition (CON). Table 50 indicates no significant main effects or interactions. Figure 57 shows that in this comparison, there is virtually no difference between the judgment type means. With respect to conditions, on the other hand, Figure 58 shows that, again, perception of events in the resource (fuel) management task was somewhat lower while that task was automated.

Level I (Perception) Questions about System Monitoring

Table 51 displays a summary analysis for Level I (Perception) questions about the system monitoring task. The two conditions entered in this analysis were automated resource management (ARM) and automated system monitoring (ASM). Table 51 shows a significant main effect for condition, **F**(1, 14) = 5.12, p < .05) and a significant interaction between judgment type and condition , **F**(1, 14) = 5.12, p <. 05). Figure 59 shows that subjects correctly answered significantly more questions about their perception of events in the system monitoring task while the resource management task was automated, as opposed to while the system monitoring task was automated, as was predicted in Experiment II, Hypothesis # 2 (pg. 34). However, Figure 60, which shows the means for the interaction between judgment type and condition shows this difference occurs only under the absolute judgment type. Under the comparative judgment type subject perception of events in the system monitoring task was identical whether resource (fuel) management or system monitoring was automated. Interestingly, one will also notice in Figure 60 that perception SA was better, overall, under the absolute judgment type, which was not predicted. Therefore, it appears that subjects were paying more attention, overall, to the system

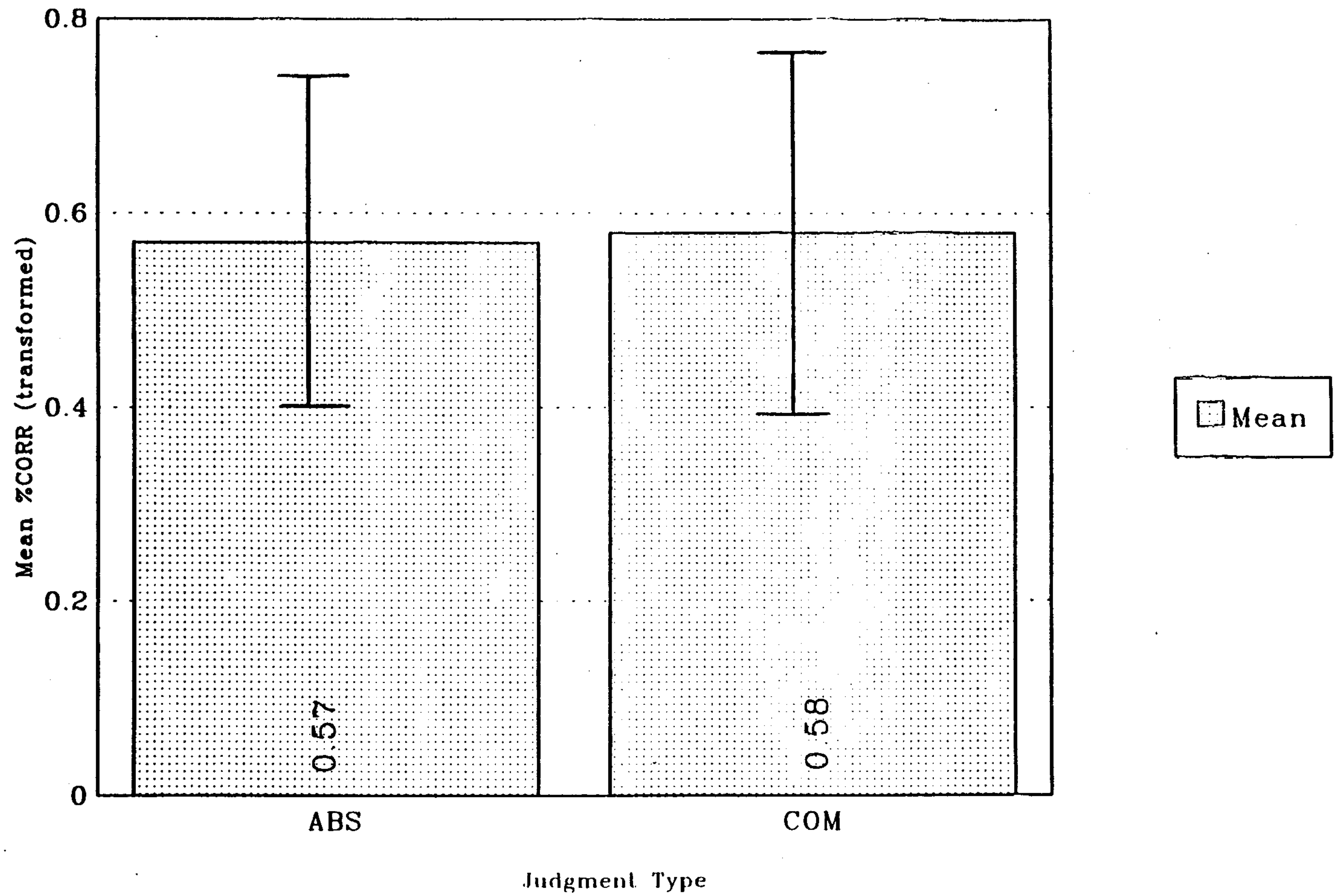


Figure 57. Judgment Type Means
Level I (Perception) SA of Resource (Fuel) Management
ARM vs CON

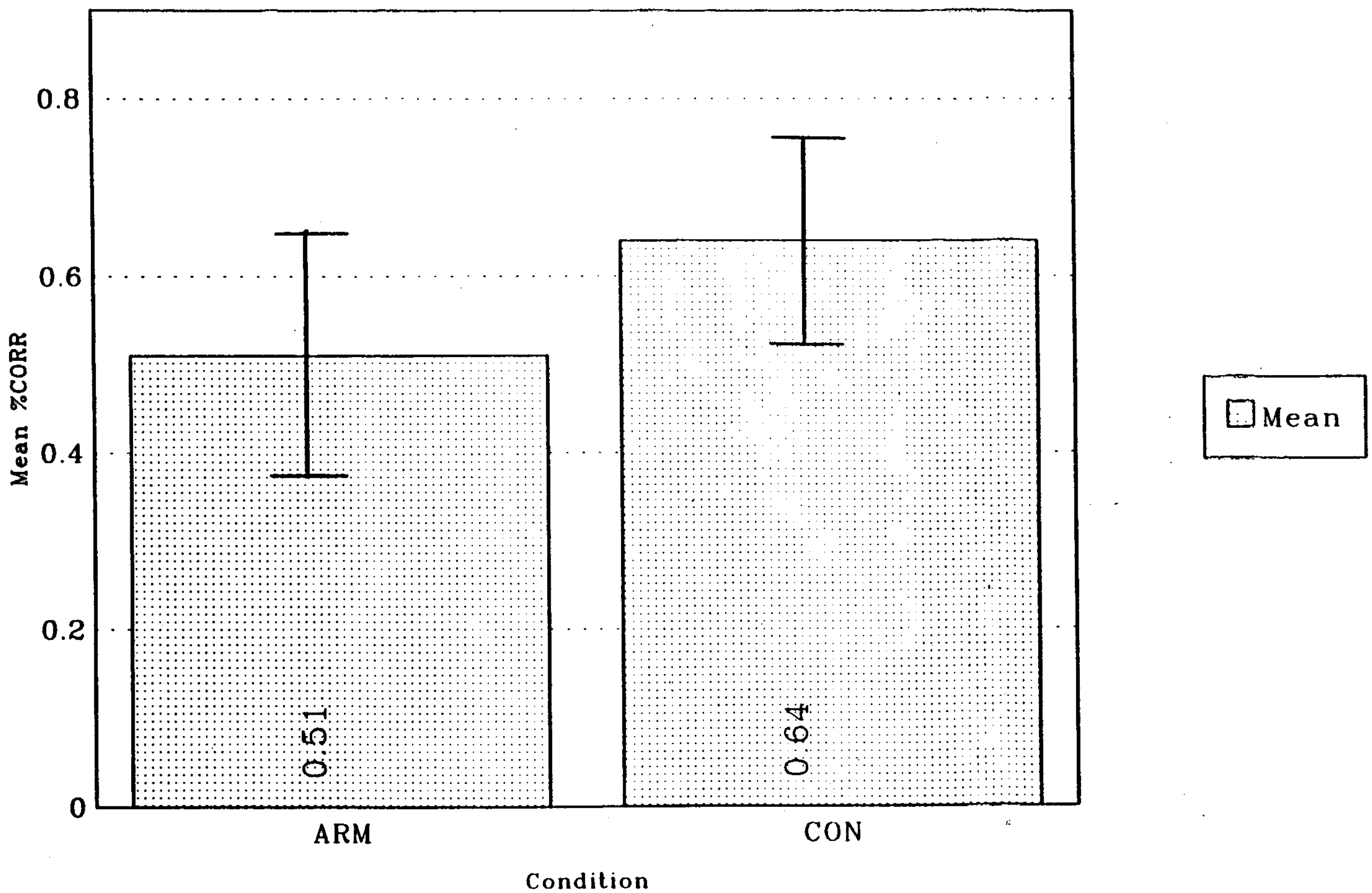


Figure 58 Condition Means
Level I (Perception) SA of Resource (Fuel) Management
ARM vs CON

Table 51: Summary Analysis for Level I (Perception) Questions about System Monitoring: ASM versus ARM

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	1.49	14	0.58	2.587	0.130
Cond	1	0.38	14	0.07	5.116	0.040
JTxCond	1	0.38	14	0.07	5.116	0.040

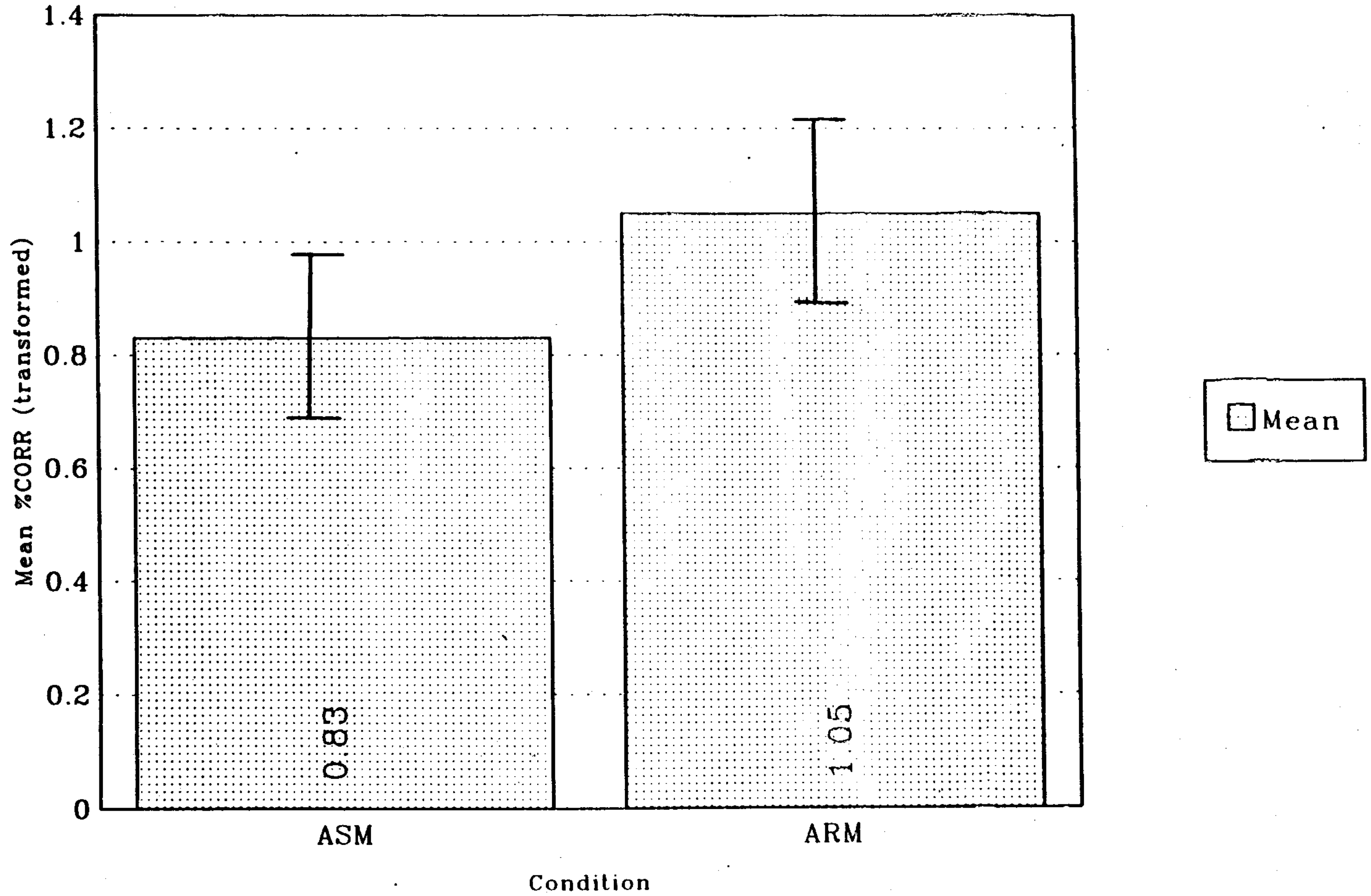


Figure 59: Condition Means
Level I (Perception) SA of System Monitoring
ASM vs ARM

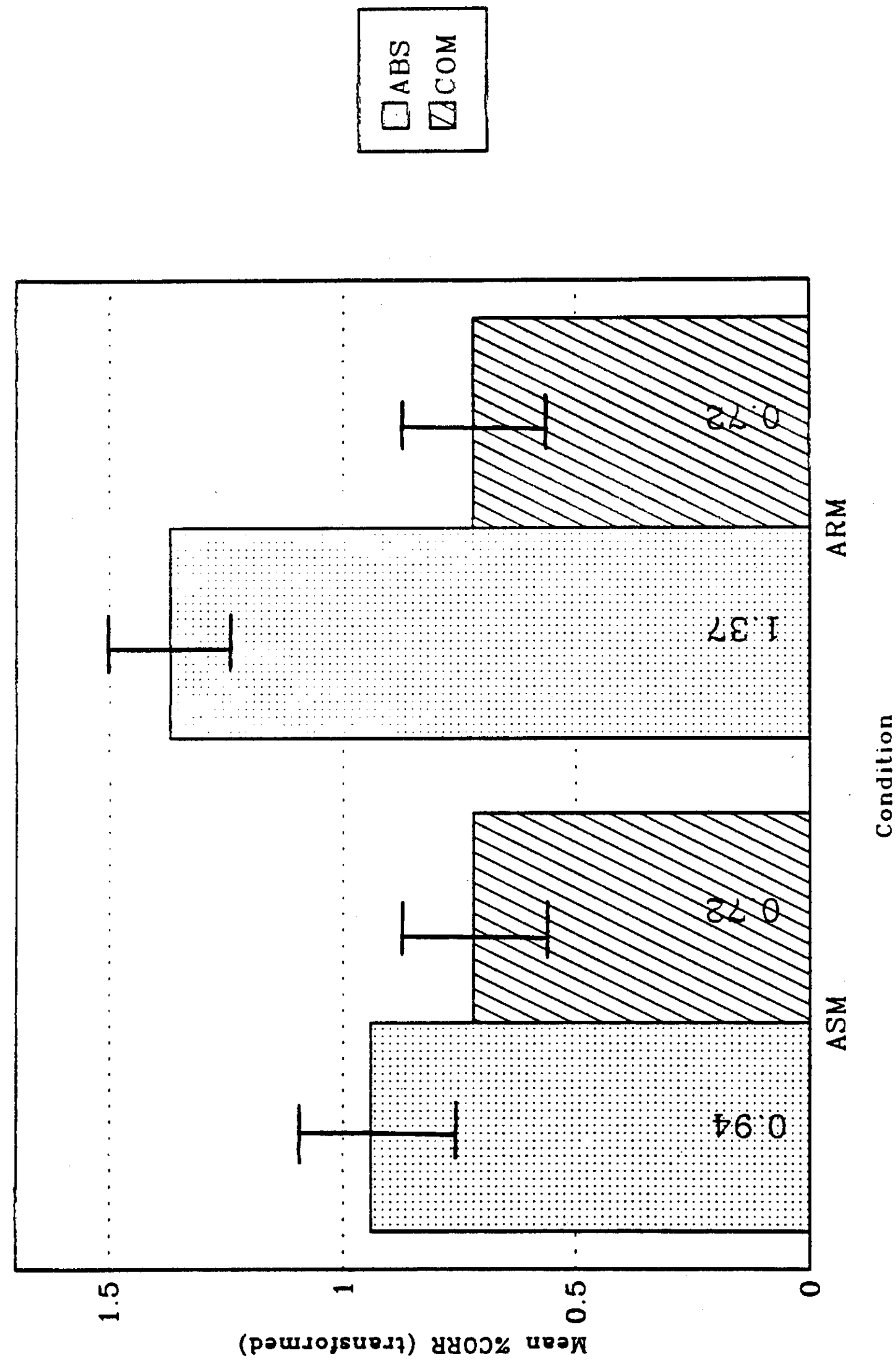


Figure 60: Judgment Type x Condition Means
Level 1 (Perception) SA of System Monitoring
ASM vs ARM

monitoring task under the absolute judgment type. Perhaps the subjects felt the absolute judgment type system monitoring task was more difficult than the comparative judgment type and therefore warranted more attention. In the case of the comparative judgment type, on the other hand, subjects paid less attention, overall. The identical means perhaps represent a floor effect.

Table 52 displays a summary analysis for Level I (Perception) questions about the system monitoring task. The two conditions entered in this analysis were automated system monitoring (ASM) and the full manual control condition (CON). Table 52 indicates there were no significant main effects or interactions in this analysis. Note, however, that Figure 61 shows the same general trend, which is contrary to expectations, of subjects showing higher perception of events in the system monitoring task under the absolute judgment type. Furthermore, Figure 62 shows the same general trend, which follows Experiment II, Hypothesis # 2 (p. 34), of lower perception of events in the system monitoring task while system monitoring was automated.

Level II (Meaning) SA Questions about Resource (Fuel) Management

Table 53 displays a summary analysis for Level II (Meaning) questions about the resource (fuel) management task. The two conditions entered in this analysis were automated resource (fuel) management (ARM) and automated system monitoring (ASM). Table 53 indicates there were no significant main effects or interactions in this analysis. However, Figure 63 shows that, as predicted in Experiment II,

Table 52: Summary Analysis for Level I (Perception) Questions about System Monitoring: ASM versus CON

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.64	14	0.31	2.074	0.172
Cond	1	0.07	14	0.25	0.279	0.606
JTxCond	1	0.38	14	0.25	0.150	0.704

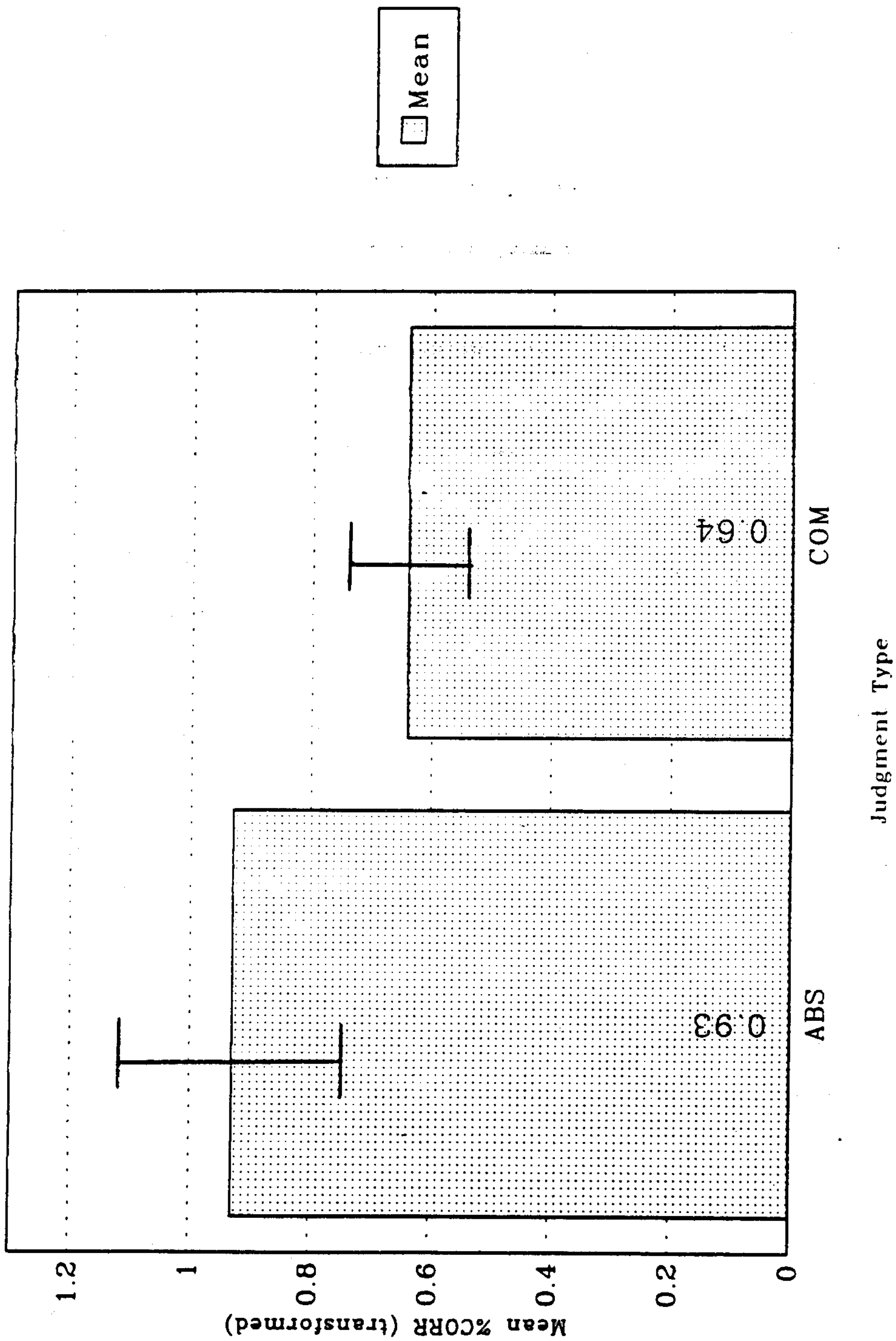
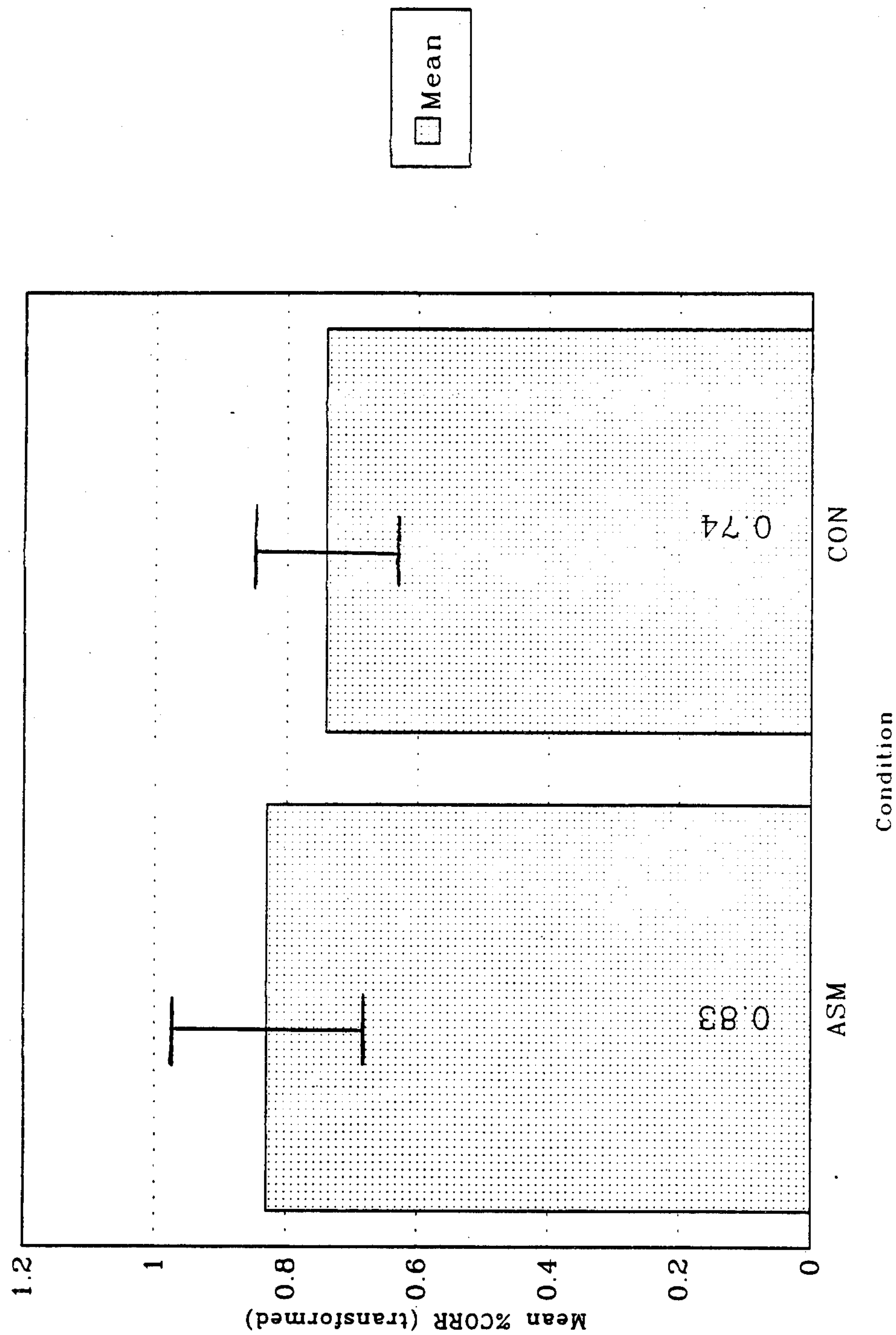


Figure 61. Judgment Type Means
Level 1 (Perception) SA of System Monitoring
ASM vs CON



...on Means
Level 1 (Perception) SA of System Monitoring
ASM vs CON

Table 53: Summary Analysis for Level II (Meaning) Questions about Resource (Fuel) Management: ARM versus ASM

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.21	14	0.102	2.082	0.171
Cond	1	0.66	14	0.177	3.725	0.074
JTxCond	1	0.18	14	0.177	1.046	0.324

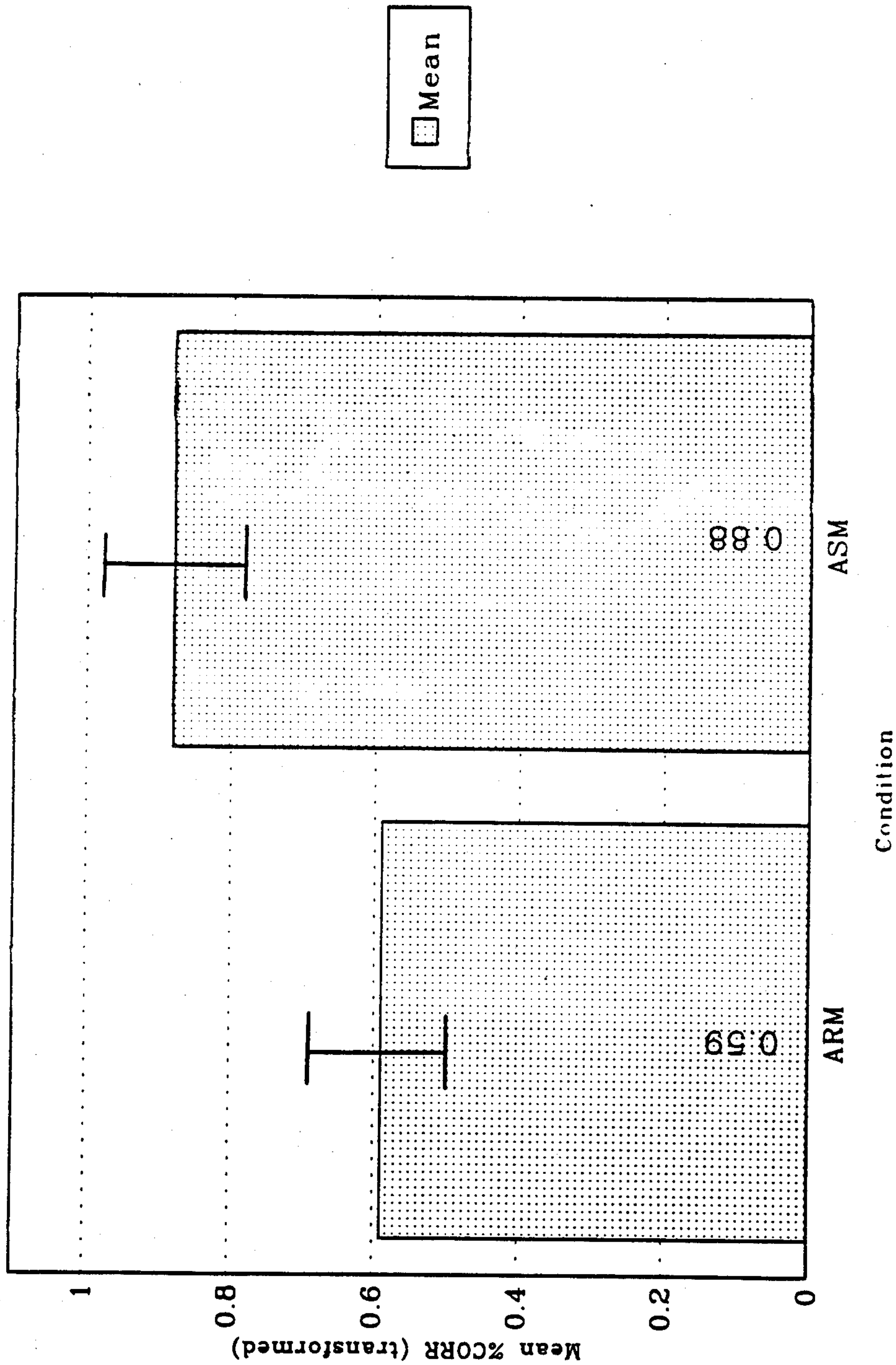


Figure 63: Condition Means
Level II (Meaning) SA of Resource (Fuel) Management
ARM vs ASM

Hypothesis # 4 (pg. 35), subjects achieved lower scores on Level II (Meaning) questions about the resource (fuel) management task while that task was automated, relative to while the system monitoring task was automated. Furthermore, as is shown in Table 54, when the condition of automated resource (fuel) management (ARM) was compared to the full manual control condition (CON), the main effect for condition approached significance, $\underline{F}(1, 14) = 4.18$, $p = .060$. Figure 64 shows that, as predicted in Experiment II, Hypothesis # 4 (pg. 35), subjects again achieved lower scores on Level II (Meaning) questions about the resource (fuel) management task while that task was automated, relative to the full manual control condition.

Level II (Meaning) SA Questions about System Monitoring

Table 55 displays a summary analysis for Level II (Meaning) questions about system monitoring. The two conditions entered in this analysis were automated system monitoring (ASM) and automated resource (fuel) management (ARM). Table 55 indicates a significant two-way interaction between judgment type and condition, $\underline{F} = 4.65$, $p < .049$. As shown in Figure 65 subjects answered significantly more Level II (Meaning) questions about system monitoring while resource (fuel) management was automated, as opposed to while system monitoring was automated, but only under absolute judgment. This follows Experiment II, Hypothesis # 3 (p. 34), which predicted there would be a difference in the number of correct Level II questions about system monitoring only under absolute judgment. Table 56 displays a summary analysis for Level II (Meaning) questions about system monitoring. There were no significant main effects or interactions.

Table 54: Summary Analysis for Level II (Meaning) Questions about Resource (Fuel) Management: ARM versus CON

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.11	14	0.18	0.572	0.462
Cond	1	0.46	14	0.11	4.177	0.060
JTxCond	1	0.86	14	0.11	0.784	0.391

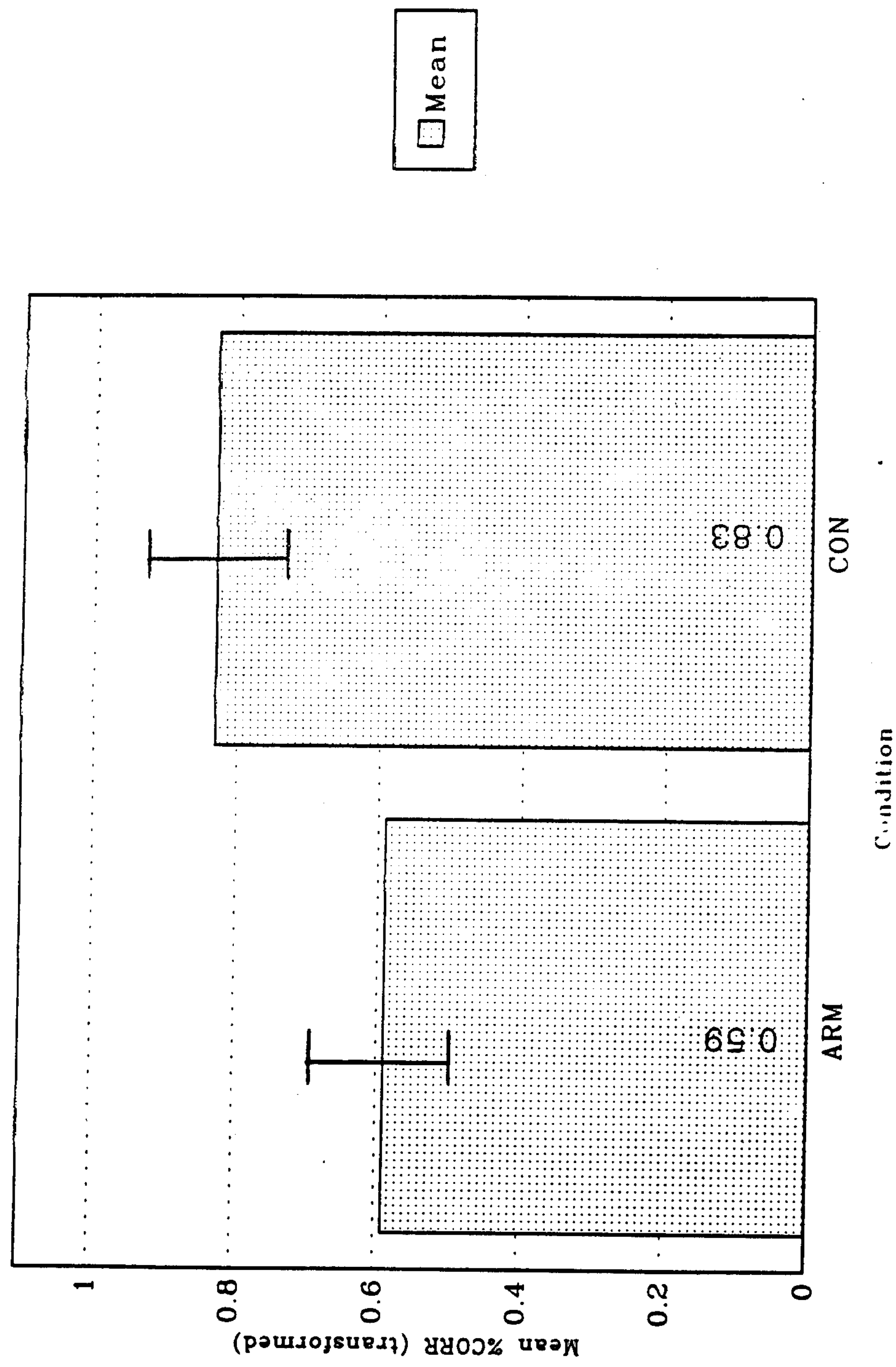


Figure 64: Condition Means
Level II (Meaning) SA of Resource (Fuel) Management
ARM vs CON

Table 55: Summary Analysis for Level II (Meaning) Questions about System Monitoring: ASM versus ARM

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.006	14	0.078	0.083	0.777
Cond	1	0.082	14	0.027	3.094	0.100
JTxCond	1	0.124	14	0.027	4.653	0.049

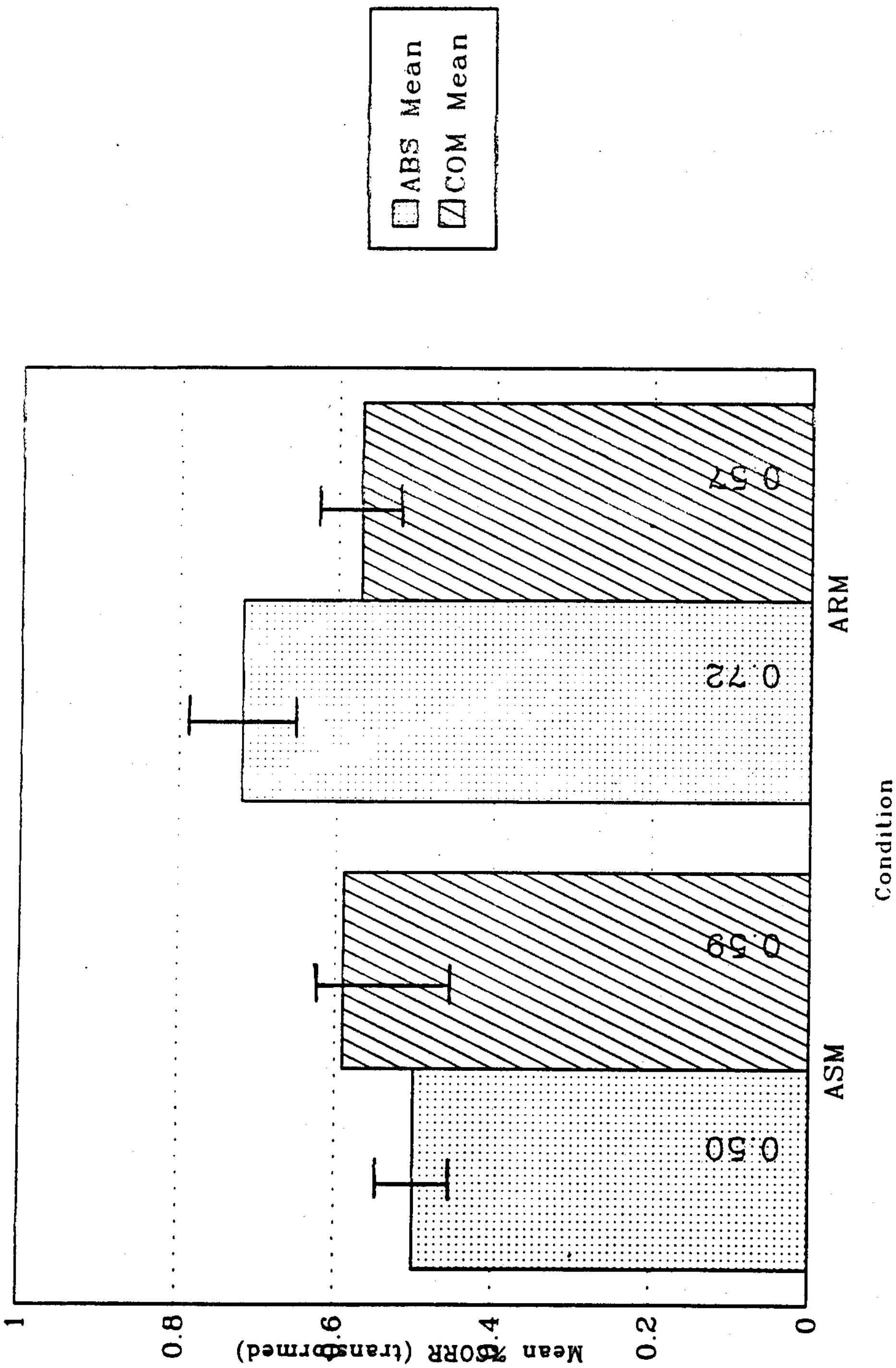


Figure 65: Condition Means
Level II (Meaning) SA of System Monitoring
ASM vs ARM

Table 56: Summary Analysis for Level II (Meaning) Questions about System Monitoring: ASM versus CON

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
JT	1	0.07	14	0.05	1.381	0.259
Cond	1	0.03	14	0.05	0.549	0.471
JTxCond	1	0.00	14	0.05	0.000	0.991

CHAPTER IV

DISCUSSION

Modern aviation, more so than ever, is a technological proving ground. The stressors encountered within the aerospace environment are such that neither man nor machine may escape them entirely; the intricate balance of these two basic elements must, therefore, be at all times at the core of aviation systems designs.

In examining the relationship between the aviator and the automated system, one must take care to understand the nature of the beast before embarking on any attempt to tame it. It is the task of the human factors designer to investigate the nature of the pilot-automation system. Because the human aviator is, presently and within the foreseeable future, the decisive controller of the aircraft, the current investigation examined his/her information processing capacities and limitations within the environment of complex, semi-automated aviation.

The problem was approached by first dissecting the components of the foundation of human information processing: decision-making. A review of the literature established four stages of this process, termed in this paper as detection, diagnosis, decision, and execution (Barrett and Donnell, 1989; Flathers, Giffin, and Rockwell, 1982; Harris et al., 1991; Wickens and Flach, 1988). A theoretical model was created to guide hypotheses with respect to how the human decision-making process operates within a complex, semi-automated environment. As it was determined that the cornerstone of human information processing within a particular situation includes one or more modifiable internal cognitive models, the role of such models within the stages of decision-making was of interest.

The internal model is defined in this paper as "learned associative relationships." Through these learned associative relationships, the human processor directs decision-making. The definition of the internal model, in addition to the evidence, signifies its ability to change as a result of interacting with the demands of a situation. If the internal model is modifiable, one must expect that the decision-making process directed by the internal model is modifiable, and that, ultimately, the outcome of the decision-making process is modifiable. Furthermore, if one were to optimize the internal model to accurately represent the situation in question, one would expect optimization of the decision-making process and its subsequent product, performance. The closer one's internal model is to accurate representation of the situation, the greater the "situational awareness" of the decision-maker.

The fact that the internal model has the ability to change as a result of interaction with its situational environment does not mean that it necessarily will. Furthermore, if a change does result from the interaction, it may not necessarily be an accurate, or even adequate, representation of the situation. The internal model determines for the decision-maker how much time should be taken before concluding the process, the amount and type of information to be used, as well as how it is to be manipulated in the decision. Such elements are affected by environmental and task factors such as level of arousal (Bahrick, Fitts, and Rankin, 1952; Bursill, 1958; Easterbrook, 1958) and level of complexity (Baddeley, 1972; Keinan, 1987; Weltman et al., 1971; Wright, 1974). The components of the internal model and their use in the decision-making process are also affected by facets of the decision-maker, such as individual differences (Gopher, 1982; Gopher and Kahneman, 1971) and experience (Braune and Trollip, 1982, Chase and Simon, 1973; Chechile et al., 1989; Gopher, 1982).

Within the vigilance domain, there is abundant evidence that the cognitive nature of the tasks affects performance outcomes (Gluckman, Warm, and Dember, 1987; Parasuraman, 1976; Parasuraman and Davies, 1977; Parasuraman, Warm, and Dember, 1987; Warm et al., 1984). In fact, vigilance researchers Fisk and Scerbo (1987), in discussing the theory of automatic versus controlled processing (Schneider and Shiffrin, 1977), indicate that properties of the task being performed are equally important to the degree of practice time in establishing automatic processing. Specifically, those tasks in which the stimuli and responses are consistently mapped (requiring the same overt responses to the same stimuli over time) were shown to produce automatic processing more readily than those tasks in which the stimuli and responses possess varied mapping (requiring subject responses to change to the same stimuli across time).

A relatively new challenge to the study of human monitoring behavior is that of automating certain tasks within a multi-task situation. Although originally viewed as a means of reducing workload and enhancing performance, automation designed without concern for the human operator as an element within the system has lead to automation-induced problems. Within the aviation world, these include potentially deadly factors, such as pilot complacency, lack of situational awareness, loss of manual proficiency, etc (Weiner, 1977; Weiner and Curry, 1980). As these problems are reminiscent of those uncovered through decades of vigilance work, many investigators have voiced concerns that over-automation in the cockpit places the pilot within the non-optimal position of "supervisory monitor." On the other hand, few would dispute the fact that the cockpit environment has become so complex, some degree of automation is an absolute necessity to prevent complete information overload of the aviation crew.

What is left, then, is a situation in which the human factors designer must optimize the effects of the specific environment (semi-automated aviation) on the cognitive nature of the specific tasks.

The present study proposed that an understanding of a basic aspect of the internal model associated with a particular task could aid in the prediction of the reaction of the internal model and subsequent decision-making performance within the context of certain environmental factors. Specifically, it was proposed that the internal model associated with certain tasks may be classified on the basis of its stability, and that this classification could be used to predict how the environment of a semi-automated, aviation-relevant, multi-task situation would differentially affect the performance of these tasks.

A task with an associated stable internal model would be one in which the information relevant to decision-making (the relationships and properties within the task), once learned, does not change across time. A task with an associated dynamic internal model, on the other hand, would be one in which the information relevant to decision-making, once learned, does change across time.

It was expected that, in contrast to a stable model task, long-cycled automation of a dynamic model task would reduce the overall cognitive demands of the multi-task environment during the period of automation, but that upon return to manual conditions, it would have detrimental effects. Those detrimental effects would be due, fundamentally, to an inability of the human operator to adequately monitor the automated task. In the case of a dynamic model task, this would then lead to a failure to continuously update the internal model guiding decision-making performance on the task. This, in turn, would equate to a loss of deeper levels of situational awareness, particularly Level II (Endsley, 1990), or the meaning of the information. The human operator of the dynamic model task, upon returning to manual performance, would be operating on the basis of an outdated internal model, while in the process of attempting to re-update that model in the potentially stressing context of post-automation failure.

Subjects were given an aviation-relevant, 3-task battery. Two of the tasks were labeled as having an associated stable internal model (system monitoring task) and an associated dynamic internal model (resource [fuel] management task), and were experimentally manipulated. Subjects underwent sessions in which all tasks were performed manually during the entire session, as well as sessions in which one of the two manipulated tasks was automated for approximately 20 minutes, after which time the automation extinguished without notification of the subject. Although one could argue that failing the automation at approximately the same moment across experimental conditions could lead the subject to anticipate the failure, it was justified on the basis that approximate equality of time intervals in exposure to automated versus manual performance was important to good measurement. Additionally, prior research in similar domains indicated the relative inability of subjects to accurately judge the passage of time when involved in such a multi-task situation (Morrison et al., 1991).

As the tasks were chosen for their representativeness to the aviation environment, as well as their clear distinction on the basis of a stable versus dynamic internal model (System Monitoring vs. Resource [Fuel] Management), several manipulations were added in order to clarify any differences between the tasks that were expected to be found in the results.

The Level of Engagement was defined as the number of signals (System Monitoring Task) or pump faults (Resource [Fuel] Management Task) per unit time. There were two levels; high (12 per 10-minute interval) and low (6 per 10-minute interval). This manipulation varied the level of interaction of the subject with the task necessary to perform it adequately. It was examined <u>within</u> tasks in order that any contribution of this variable to differences in the effects of automation <u>between</u> tasks could be examined separately, and in conjunction with, the theorized model distinctions.

Judgment type (absolute vs. comparative) was manipulated in order to examine the stable versus dynamic model concept <u>within</u>, as well as between tasks. The present study was designed to examine the theory that aviation-relevant tasks are affected by automation and automation failure in a manner which can be predicted on the basis of the stability of the internal model associated with the task.

The theory held that tasks associated with a stable model would be more heavily weighted on the detection stage of decision-making. By definition, a stable model is one in which the information relevant to decision-making, once learned, does not change across time. The crucial aspect of the subject decision would then be the actual detection of a potential problem (i.e., an out-of-range signal). Once detected, the remaining decision-making stages could be guided by the stable model. Also because of the stable nature of the model associated with a particular task, any loss in situational awareness (due to factors such as long cycles of automation) would be expected to affect primarily those levels of situational awareness crucial to detection. Specifically, when applying the Endsley (1991) levels of situational awareness, the theory would predict that Level I, perception of an event, would be most affected in terms of a loss of situational awareness. It would not be expected that one would lose significant awareness of the meaning (Level II) or future projection (Level III) of an event in the case of a stable model, as these things do not change across time.

On the other hand, the theory held that tasks associated with a dynamic internal model would be more heavily weighted on the diagnosis stage of decision-making. The subject would need to constantly reassess the situation and possibly restructure the internal model and associated decisions/responses. Related to this dynamic nature, the theory predicted that any losses in situational awareness would be manifest at deeper levels (II & III) than in the case of a more stable model. The meaning of the information and its future projection would change across time.

These theorized differences were used to characterize the system monitoring and resource (fuel) management tasks in the present study. Their application to the theoretical model of decision-making in a semi-automated cockpit formed the basis for several predictions regarding performance, workload and situational awareness.

In dealing with a multi-task situation, one would expect a more dynamic internal model to have the potential to be more inaccurate. This would be expected on the basis that any loss of situational awareness reduces the validity of the information in the dynamic internal model. A loss of situational awareness in a particular task within a multi-task situation may occur either from distraction from other tasks (attention drawn away from task in question) or from long cycled automation of the task in question.

While the task in question was not automated, one would expect the more dynamic model task to be more attention-demanding, because of the increased need to maintain situational awareness in order to maintain model accuracy (and consequently, decision accuracy). While that same task was automated, however, one would expect that situational awareness would be lost at deeper levels (II and III) than in the case of a task associated with a more stable model. Performance on the former task would then be expected to be more detrimentally affected with respect to regaining manual control following automation.

Although some of the specific predictions regarding task type were not realized, there was much evidence in support of the predictions regarding differences in task type.

First, performance on the tracking task was enhanced during periods when the resource (fuel) management task was automated, but not during periods when the system monitoring task was automated. Although it was originally predicted that tracking performance would be enhanced in either case, the actual finding does support the general prediction of task differences. The fact that tracking performance was improved during automation of the resource (fuel) management task but not during automation of the system monitoring task supports the expectation that a task with a more dynamic internal model would demand more attention.

Second, as predicted, performance in both the system monitoring and the resource (fuel) management tasks was enhanced during automation of the opposing tasks, but in ways which differed as predicted. Although performance of the resource (fuel) management task was enhanced by automation of the system monitoring task, there was no significant difference between absolute and comparative judgment types. In the case of the system monitoring task, on the other hand, while the detection performance was enhanced by automation of the resource (fuel) management task, it was also differentially affected by judgment type, supporting the prediction that the absolute judgment type would add instability to the stable system monitoring model.

Third, as predicted, performance in block 5 of the resource (fuel) management task was worse following automation of that task (ARMHH and ARMLH) than in all other conditions. Furthermore, a close examination of the data reveals some interesting facets. In blocks 5 and 6, performance on the resource (fuel) management task was worse for ARMHH than for ARMLH (although not significantly). There are three potential explanations for this finding. First, it is possible that the subject experienced increased workload with the higher engagement level of the resource (fuel) management task. This, however, is not supported by the NASA-TLX data. A second means of accounting for the difference could be an increased chance for error in the high as opposed to low engagement level. This, too, is not supported by the data. In blocks 1-4, performance was significantly worst in the control condition in which the engagement level of the resource (fuel) management task was low. Additionally, there were no significant differences in performance on the resource (fuel) management task in any of the controls of blocks 5 or 6. A third possibility is that it is easier for the subject to "catch up" upon losing control of the resource (fuel) management task when it is of a higher engagement level. One would then expect this discrepancy to become more evident following automation failure, as the subject must reorient him/herself to the overall challenge of performing three tasks simultaneously, as well as to the resource (fuel) management task itself. Furthermore, the evidence from block 6, in which conditions ARMHH and ASMHH were significantly worse than all others, indicates that subjects' attempts to readapt to performing all three tasks following <u>any</u> automation failure have continuing detrimental effects on performance of the resource (fuel) management task, when it is of a high engagement level. Again, this supports the supposition that resource (fuel) management performance detriments resulting from reorienting to performance of three tasks simultaneously would be more evident under higher levels of engagement of this task, as the higher level demands greater attention and interaction.

As the performance data with respect to the resource (fuel) management data is somewhat complicated, the most compelling evidence in support of hypotheses regarding separating tasks on the basis of a dynamic versus stable internal model is provided by the SAGAT data. The percentage of Level II (Meaning) questions about the resource (fuel) management task that were answered correctly dropped when the resource (fuel) management task was automated. On the other hand, subjects were able to answer Level II questions about the system monitoring task regardless of whether or not it was automated, except under the absolute judgment type. This supports the stability of the model associated with the system monitoring task and the instability added by an absolute judgment requirement.

In the case of the system monitoring task, both false alarm and detection performance were significantly affected by automation. Specifically, performance on the system monitoring task was detrimentally affected following both ASMHH and ASMHL when compared to automation of the resource (fuel) management task, or to the full manual controls, although this finding was predicted only for low levels of engagement.

Additionally, the block 5 false alarm rate for ASMHH was significantly greater than for ARMHH or ASMHL, but only under Absolute Judgment Types. Furthermore, although the false alarm rate was greater for Absolute ASMHH than for Absolute

ARMHH, recall that the detection rate was significantly lower, overall, for Absolute Judgment Types. Additionally, the detection rate was significantly lower, overall, for ASMHH as compared to ARMHH. This is interesting because typically, one would expect to see a greater detection rate with a greater false alarm rate, indicating a more lenient response bias. A plausible explanation would be that the level of expectation under the higher engagement level, coupled with a less stable, more uncertain model (due to absolute judgment type), would lead the subject to shift to a more lenient response bias (accounting for increased false alarm rate), but one based upon an internal model which, due to greater instability, has increased possibility of inaccuracy. In other words, the subjects responded more frequently on the basis of a more inaccurate model.

It would appear, then, that automation of the system monitoring task does detrimentally affect the ability to return to manual operation. However, the interaction between judgment type and condition with respect to block 5 false alarm rate lends support to hypotheses regarding the stability of the internal model associated with a task. This is particularly true when one considers the fact that there were no significant differences (considering block 5 false alarm rate) among comparable full manual control conditions between judgment types (i.e., absolute CONHH compared with comparative CONHH). It was only following automation of the system monitoring task that the differences in the judgement types had a significant effect on the false alarm rate. Finally, it is important to note that throughout all blocks, percent correct was significantly affected by judgment type, to the degree that the absolute judgment type lead to consistently worse performance on this dependent variable. Although one might argue this effect was due to decreased sensitivity for the system monitoring task under absolute judgment conditions, an initial pilot study found no significant difference in any of the three dependent variables between absolute and comparative conditions (conditions in the pilot study involved 5-minute, full manual control of the system monitoring task alone). These pilot study results thus indicated at least a floor effect in terms of sensitivity, and thereby allowed for any further results to be interpreted on a more cognitive level.

CHAPTER V

SUMMARY

In summary, one would have to conclude the findings with respect to the performance data do not unequivocally support a stable versus dynamic internal model task distinction as the basis for differential affects of automation and return-to-manual-operation on various flight-related tasks. However, the performance data definitely indicate general task-dependent effects of automation and return-to-manual-operation. Furthermore, the facts do not dispute task distinction on the basis of the stability of the internal model. In fact, evidence from the SAGAT data indicates significant differences in situational awareness following automation of the two tasks, in such a manner as was predicted on the basis of a stable versus dynamic internal model. Additionally, there is enough evidence in the overall findings to support continued study in this domain. In fact, preliminary results from a study conducted by Gluckman, Carmody, Morrison, Hitchcock, and Warm (in progress), indicate strong task-dependent effects under short (10 minute) cycles of automation, and various automation strategies. These effects strongly support the notion that not only automation, but automation strategy differentially affect performance in a manner predictable on the basis of the stability of the internal model associated with a task.

On the basis of the results in the present study, it was determined that there is strong evidence for distinguishing tasks on the basis of the stability of the internal cognitive model and how performance may be affected by long-cycled static automation and automation failure. Although the results were admittedly not conclusive on this point, the study was designed to be exploratory in nature; to establish a basis for investing further, higher fidelity resources into this experimental vein. It was intended to examine the potential for classifying specific tasks in the overall aviation battery in terms of some definable cognitive type, as well as the potential for predicting how certain automation designs would affect said task types. This is an important avenue of research if technology is to refine cockpit automation designs of the future with the goal of optimized efficiency, safety, and performance.

Future research within this domain should concentrate upon refined measurements, more homogenous groups (pilots of same level and type experience), and more realistic, higher fidelity simulations. Additionally, potential training benefits could be gleaned from examinations into levels of experience (novice vs. expert pilots) and the manner in which internal models change and are executed.

REFERENCES

Atkinson, R.C. and Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press.

Babbit, B.A., and Nystrom, C.O. (1985). *Training and Human Factors Research on Military Systems: Questionnaire Construction Manual (Army Manual P-71-1)*, Fort Hood, TX: Army Research Institute for the Behavioral Sciences.

Bacon, S.J. (1974). Arousal and the range of cue utilization. *Journal of Experimental Psychology*, 102(1), 81-87.

Baddeley, A.D. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology*, 63 (4), 537-546.

Bahrick, H.P., Fitts, P.M., and Rankin, R.E. (1952). Effect of incentives upon reactions to peripheral stimuli. *Journal of Experimental Psychology*, 44, 400-406.

Barrett, C.L. and Donnell, M.L. (1990). Real time expert advisory systems: Considerations and imperatives. *Information and Decision Technologies*, 16, 15-25.

Becker, A.B., Gluckman, J.P., Warm, W.N., Dember, W.N., and Hancock, P.A.(1987). *Subjective Workload in Dual-Task Monitoring*, Paper presented at the Annual Meeting of the International Psychonomics Society.

Bergeron, H.P. and Hinton, B.S. (1985). Aircraft automation: The problem of the pilot interface. *Aviation, Space, and Environmental Medicine*, Feb, 144-148.

Boff, K.R., Kaufman, L., and Thomas, J.P. (Eds). (1986). *Handbook of Perception and Human Performance, Volume II: Cognitive Processes and Performance*, New York: John Wiley and Sons.

Braune, R.J. and Trollip, S.R. (1982). Towards an internal model in pilot training. *Aviation, Space, and Environmental Medicine*, 53(10), 996-999.

Buch, G. (1984). An investigation of the effectiveness of pilot judgment training. *Human Factors*, 26(5), 557-564.

Bursill, A.E. (1958). The restriction of peripheral vision during exposure to hot and humid conditions. *Quarterly Journal of Experimental Psychology*, 10, 113-129.

Chambers, A.B. and Nagel, D.C. (1985). Pilots of the future: Human or computer? *Communications of the Association for Computing Machinery*, 28, 1187-1199.

Chase, W.G. and Simon, H.A. (1973). The mind's eye in chess. *Visual Information Processing*, 215-281, San Diego: Academic Press, Inc.

Chechile, R.A., Eggleston, R.G., Fleischman, R.N., and Sasseville, A.M. (1989). Modeling the cognitive content of displays. *Human Factors*, 31(1), 31-43.

Clifton, JR., C. and Slowiaczek, M.L. (1981). Integrating new information with old knowledge. *Memory and Cognition*, 9(2), 142-148.

Collins and E.E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 461-494). San Mateo: Morgan Kaufmann Publishers, Inc.

Comstock, J.R. and Arnegard, R.J. (1990). Multi-attribute task battery. NASA Langley Research Center, Hampton, VA. Report in progress.

Corcoran, D.W.J., Dennett, J.L., and Carpenter, A. (1970). Cooperation of listener and computer in a recognition task: Effects of listener reliability. *Journal of the Acoustical Society of America*, 47, 480-488.

Davis, D.R. (1948). Pilot error. H.M. Stationary Office, London England. Air Publication no. 3139A.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 60(3), 183-201.

Edwards, W. and Slovic, P. (1968) Seeking information to reduce the risk of decisions. *Human Factors*, 4, 187-197.

Einhorn, H.J., Kleinmuntz, D.N, and Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgement. *Psychological Review*, 86, 465-485.

Einhorn, H.J. and Hogarth, R.M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology*, 32, 53-88.

Endsley, M. R. (1990). *Situation Awareness Global Assessment Technique (SAGAT): Air to Air Tactical Version: User's Guide*. Northrop Corporation, Hawthorne, CA. NOR DOC no. 89-58 REV A.

Endsley, M.R. (1991). Situation Awareness in Dynamic Human Decision Making: Measurement. Department of Industrial Engineering, Texas Tech University, Lubbock, TX. Unpublished document.

Endsley, M.R. (1991). Situation Awareness in Dynamic Human Decision Making: Theory. Department of Industrial Engineering, Texas Tech University, Lubbock, TX. Unpublished document.

Fisk, A.D. and Scerbo, M.W. (1987). Automatic and control processing approach to interpreting vigilance performance: A review and reevaluation. *Human Factors*, 29(6), 653- 660.

Fisk, A.D. and Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors*, 23(6), 737-750.

Flathers, G.W., Giffin, W.C., and Rockwell, T.H. (1982). A study of decision-making behavior of aircraft pilots deviating from a planned flight. *Aviation Space and Environmental Medicine*, 53(10), 958-963.

Gluckman, J.P. (1990). Changing task demands in sustained attention: Effects on performance and perceived workload. Unpublished doctoral dissertation, University of Cincinnati, Cincinnati, OH.

Gluckman, J.P., Morrison, J.G., and Deaton, J. (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. *Proceedings of the Human Factors Society 35th Annual Meeting*, San Francisco, CA, 116-120.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. *Human Factors*, 24(2), 173-183.

Gopher D., and Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills*, 33, 1335-1342.

Green, D.M. and Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Greenhouse, S.W. and Geisser, S. (1958). Extension of Box's results on the use of the F distribution in Multivariate analysis. *Annals of Mathematical Statistics*, 29, 95-112.

Grossberg, S. and Gutowski, W. (1987). Neural dynamics of decision making under risk: Affective balance and cognitive-emotional interactions. *Psychological Review*, 94(3), 300-318.

Harris, S.D., Owens, J., Barrett, C.L., Parisi, M., and Becker, D. (1991). Sensor fusion and situation assessment: A perspective. Code AIR-05TP3, Naval Air Systems Command, Washington, D.C. Unpublished document.

Hart, S.G. and Bortolussi, M.R. (1984). Pilot errors as a source of workload. *Human Factors*, 26(5), 545-556.

Hart, S.G., Childress, M.E., and Bortolussi, M.R. (1981). Defining the subjective experience of workload. *Proceedings of the Human Factors Society 25th Annual Meeting*, 527-532.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshtaki (Eds), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hayes-Roth, B. (1977). Evolution of cognitive structures and processes. *Psychological Review*, 84(3), 260-278.

Hollnagel, E. (1987). Information and reasoning in intelligent decision support systems. *International Journal on Man-Machine Studies*, 27, 665-678.

Houston, J.P. (1986). *Fundamentals of Learning and Memory*, Orlando: Harcourt Brace Jovanovich.

Howell, D.C. (1982). *Statistical Methods for Psychology*. Boston: PWS Publishers.

Hurst, R. (Ed.). (1976). *Pilot Error: A Professional Study of Contributory Factors*, Great Britain: Granada Publishing Limited.

Ingleby, J.D. (1974). Further studies of the human observer as statistical decision maker. *Organizational Behavior and Human Performance*, 12, 299-314.

Itoh, Y., Hayashi, Y., Tsukui, I., and Saito, S. (1990). The ergonomic evaluation of eye movement and mental workload in aircraft pilots. *Ergonomics*, 33(6), 719-733.

Janis, I.L. (1982). Decision-making under stress. In L. Goldberger and S. Breznitz (Eds.), *Handbook of Stress: Theoretical and Clinical Aspects* (pp. 69-80). New York: Free Press.

Janis, I.L. and Mann, L. (1977). *Decisionmaking: A Psychological Analysis of Conflict, Choice, and Commitment.* New York: Free Press.

Jensen, R.S. and Benel, R.A. (1977). Judgement evaluation and instruction in civil pilot training (Final Report FAA-RD-78-24), National Technical Information Service, Springfield, VA.

Jerison, H.J. (1963). On the decrement function in human vigilance. In D.N. Buckner and J.J. McGrath (Eds.) *Vigilance: A Symposium* (pp. 199-212). New York: McGraw-Hill.

Kanarick, A.F., Huntington, J.M., and Petersen, R.C. (1969). Multi-source information acquisition with optional stopping. *Human Factors*, 11(4), 379-386.

Kantowitz, B.H. and Casper, P.A. (1988). Human workload in aviation. In Wiener, E.L. and Nagel, D.C. (Eds.), *Human Factors In Aviation*. San Diego, CA: Academic Press, Inc.

Keinan, G. (1987). Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology*, 52(3), 639-644.

Lester, Diahl, and Buch (1985). Private pilot judgement training in flight school settings: A demonstration project. In R.S. Jensen (Ed.), *Proceeding of the Third Symposium in Aviation Psychology*, Columbus, OH: OSU Aviation Psychology Laboratory.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6-21.

Medin, D.L. and Smith, E.E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 1984, 35, 113-138.

Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (ed.), *The Psychology of Computer Vision* (pp 211-277). New York. McGraw Hill.

Mish, F. C. (Ed.). (1986). *Webster's Ninth New Collegiate Dictionary.* Springfield, MA: Merriam-Webster.

Morrison, J.G., Gluckman, J.P., and Deaton, J.E. (1991). Adaptive function allocation for intelligent cockpits. Cockpit automation study 1: Baseline study. Air Vehicle and Crew Systems Technology Department (6021), NADC, Warminster, PA. Report no: NADC-91028-60.

Murrell, G.A. (1977). Combination of evidence in a probabilistic visual search and detection task. *Organizational Behavior and Human Performance*, 18, 3-18.

Nagel, D.C. Human error in aviation operations. In Weiner, E.L. and Nagel, D.C. (Eds.), *Human Factors In Aviation*, 1988. San Diego, CA: Academic Press, Inc.

Nelson, K. (1977). Cognitive development and the acquisition of concepts. In R.C. Anderson, R.J. Spiro, and W.E. Montague (Eds.), *Schooling and the Acquisition of Knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Neville and Moray. (1986). Monitoring behavior and supervisory control. In Boff, K.R., Kaufman, L. and Thomas, J.P. (Eds), *Handbook of Perception and Human Performance: Volume II: Cognitive Processes and Performance*. New York: John Wiley and Sons.

Parasuraman, R. (1976). *Task Classification and decision processes in monitoring behavior*, Unpublished doctoral dissertation, University of Aston, Birmingham.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29(6), 695-706.

Parasuraman, R., Bahri, T., and Molloy, R. (1991). Adaptive automation and human performance: I. Multi-task performance characteristics. (Technical Report No. CSL-N91-1), Cognitive Science Laboratory, The Catholic University of America, Washington, D.C.

Parasuraman, R. and Davies, D.R. (1976). Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 578-590.

Parasuraman, R. and Davies, D.R. (1977). A taxonomic analysis of vigilance performance.In R.R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates*, pp. 559-574. New York: Plenum.

Parasuraman, R., Warm, J.S., and Dember, W.N. (1987). Vigilance: Taxonomy and utility. In L.S. Mark, J.S. Warm, and R.L. Muston (Eds.), *Ergonomics and Human Factors: Recent Research*, 11-23. New York: Springer-Verlag.

Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G., and Barnes, M. (1990). *Theory and Design of Adaptive Automation in Aviation Systems* (Technical Report). Naval Air Development Center, Warminster, PA.

Rumelhart and Ortony (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, and W.E. Montague (eds.), Schooling and the Acquisition of Knowledge. Hillsdale, NJ: Laurence Erlbaum.

Sanders, M.S. and McCormick, E.J. (1987). *Human Factors in Engineering and Design*. New York: McGraw-Hill.

Schank, R.C. and Abelson, R.P. (1977). *Scripts, Plans, Goals, and Understanding*, chpts1-3, pp. 1-68. Hillsdale, NJ: Erlbaum and Lawrence.

Schneider, W. and Shiffrin, R.M. (1977). Control and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.

Schneider, W. and Fisk, A.D. (1982). Dual task automatic and control processing: Can it be done without cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 261-278.

Sheridan, T.B. (1987). Supervisory control. In G. Salvendy (ed.), *Handbook of Human Factors/Ergonomics*. New York: Wiley.

Simmel, E.C. and Shelton, R. (1987). The assessment of nonroutine situations by pilot two-part process. *Aviation, Space, and Environmental Medicine*, 58, 1119-1121.

Singh, Indramani, L., Molloy, R., and Parasuraman, R. (1991). *Automation-Induced "Complacency": Development of a Complacency-Potential Scale*. (Technical Report No. CSL-A-91-1), 1991, Cognitive Science Laboratory, The Catholic University of America, Washington, D.C.

Slovic, P. and Lichtenstein, S. (1971). Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior an Human Performance*, 6, 649-744.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28, 1-39.

Smith, R.P., Warm, J.S. and Alluisi, E.A. (1966). Effects of temporal uncertainty on watchkeeping performance. *Perception and Psychophysics*, 1, 293-299.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74(11), 1-29.

Statsoft (1992). Complete Statistical System (CSS): STATISTICA. Tulsa, OK.

Stone, R.B., Babcock, G.L., and Edmunds, W.W. (1985). Pilot judgment: An operational viewpoint. *Aviation, Space, and Environmental Medicine*, 56 149-152.

Tabachnick, B.G. and Fidell, L.S. (1983). *Using Multivariate Statistics*. New York: Harper & Row.

Telfer, R. (1989). Pilot decision making and judgement. In R.S. Jensen (Ed.), *Aviation Psychology*. Brookfield, VT: Gower Publishing Company.

Trejo, L.J., Lewis, G.W., and Blankenship, M.H. (1987). Brain activity during tactical decision-making: II. Probe-Evoked potentials and workload. NPRDC TN 88-12.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281-299.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 1124-1130.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 30, 453-458.

Vickers, D. and Leary, J.N. (1983). Criterion control in signal detection. *Human Factors*, 25(3), 283-296.

Vidulich, M.A. (1989). The use of judgement matrices in subjective workload assessment: The Subjective WORkload Dominance technique. *Proceedings of the Human Factors Society 33rd Annual Meeting*, pp. 1406-1410.

Vidulich, M.A. and Tsang, P.S. (1987). Absolute magnitude estimation and relative judgment approaches to subjective workload assessment. *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 1057-1061.

Ward, G.F., and Hassoun, J.A. (1990). *The Effects of Head-Up Display (HUD) Pitch Ladder Articulation, Pitch Number Location and Horizon Line Length on Unusual Attitude Recoveries for the F-16*, (Final Report ASD-TR-90-5008), Wright-Patterson AFB, OH: Crew Station Evaluation Facility, Human Factors Branch, ASD/ENECH.

Warm, J.S. and Alluisi, E.A. (1971). Influence of temporal uncertainty and sensory modality of signals on watchkeeping performance. *Journal of Experimental Psychology*, 87, 303-308.

Warm, J.S. (1990). Vigilance and target detection. In C.D. Wickens and B. Huey (Eds), Teams in transition: Workload, stress, and human factors. Washington, D.C.: National Research Council.

Warm, J.S., Howe, S.R., Fishbein, H.D., Dember, W.N. and Sprague, R.L. (1984).Cognitive demand and the vigilance decrement. In A. Mital (Ed.) *Trends in Ergonomics/Human Factors I*. North-Holland: Elsevier Science Publishers B.V.

Watchel, P.L. (1967). Conceptions of broad and narrow attention. *Psychological Bulletin*, 68(6), 417-429.

Weiner, E.L. (1977). Controlled flight into terrain accidents: System-induced error *Human Factors*, 19(2), 171-181.

Weiner, E.L. (1988). Cockpit automation. In Weiner, E.L. and Nagel, D.C (Eds), *Human Factors In Aviation*. San Diego, CA: Academic Press, Inc.

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

Weltman, G., Smith, J.E., and Egstrom, G.H. (1971). Perceptual narrowing during simulated pressure-chamber exposure. *Human Factors*, 13(2), 99-107.

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

Wickens, C.D. and Flach, J.M. (1988). Information processing. In E.L. Weiner and D.C. Nagel (Eds.) *Human Factors in Aviation*. San Diego, CA: Academic Press, Inc.

Wickens, C.D. and Kramer, A. (1985). Engineering psychology. *Annual Review of Psychology*, 36, 307-348.

Wiener, C.D. and Nagel, D.C. (Eds.). (1988). *Human Factors in Aviation*, San Diego, CA: Academic Press, Inc.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology*, 59(5), 555-561.

Wright, P. and Weitz, B. (1977). Time horizon effects on product evaluation strategies. *Journal of Marketing Research*, 14, 429-443.

APPENDIX A

REVIEW OF THE LITERATURE

Human Decision-Making in Complex, Semi-Automated Aircraft Systems

Telfer (1989), in quoting Lester, Diahl, and Buch (1985), defines pilot judgment as "the mental processes by which pilots recognize, analyze and evaluate information about themselves, their aircraft, and the operational environment, leading to a timely decision which contributes to safe flight" (p. 168).

According to several authors, pilot decision errors account for many aircraft accidents and incidents (Buch, 1984; Itoh, Hayashi, Tsukui, and Saito, 1990; Jensen and Benel, 1977; Wickens and Flach, 1988) including over half of those involving pilot fatalities (Hurst, 1976). Pilot decision-making/problem-solving is, therefore, an aspect of the aviation system that must be examined as to the extent of its role, or potential role, in total system performance. In order to do this, one must first evaluate the decision-making task within the aviation environment.

Wickens and Flach (1988) highlight three aspects of the pilot's decision-making task. First, the aviator must deal with the assessment of many and varied sources of information. Second, this assessment is based upon highly probabilistic data. Finally, the pilot must be considerate of the potential outcomes of his/her decision.

The task of decision-making within the aviation environment is, quite obviously, one involving complex, uncertain, risky decisions. Such decisions are often made under high time and load stresses, and often with the potential for producing hazardous results. Recent research within the cognitive sciences has taken an interest in human decision-making under circumstances similar to these (Edwards and Slovic, 1967; Grossberg and Gutowski, 1987; Kanarick, Huntington, and Petersen, 1969; Keinan, 1987; Tversky, 1972; Tversky and Kahneman, 1974; Tversky and Kahneman, 1981). Before venturing into such investigative tracts, however, one must first attempt to understand the fundamentals of general human information processing.

Normative versus Descriptive Work:

In the past, much of the work in the area of decision making centered about laboratory studies of normative human decision making, in which attempts were made to establish mathematical equations that could be expected to predict human decision making behaviors across situations (Einhorn and Hogarth, 1981; Slovic and Lichtenstein, 1971). Although such studies have great merit in understanding the logic underlying sequential decisions, dynamic environments often involve a different process of decision making.

According to Grossberg and Gutowski (1987),

> Normative theories are prescriptive in nature because they are concerned with devising decision-making procedures or algorithms that are optimal with regard to some set of inuitively reasonable constraints. Descriptive theories, on the other hand, are concerned with providing an accurate portrayal of how individuals actually make decisions, independent of whether those decisions are optimal or even logical (p. 300).

It has been suggested that "it is time that judgmental research became more concerned with the task conditions systematically affecting the use of different judgmental strategies rather than general model building without concern for the task environment" (Wright, 1974, p. 561). This point is perhaps particularly relevant when considering the aviation task environment, as its dynamic nature is drastically different from the static nature characteristic of tasks typically evaluated in normative pursuits.

Murrell (1977) describes dynamic tasks as involving "continuous monitoring of the information sources, or, alternatively, a series of discrete decisions each involving parallel multiple information inputs" (p. 3). Thus, the decision maker in such a task environment is forced to consider a variety of information sources and possible decision alternatives simultaneously, and select which ones are appropriate to the decision of immediate concern. Therefore, the point of focus in examining the decision process of the dynamic environment should be concerned, not so much with the underlying logic in the weighting of information, but with the actual selection of the information to be used. There is a new demand, as well as a great challenge, to uncover the psychology of decision-making environments; the myriad of components which underlay human information processing and which may, at times, defy the logic of a more structured world. This is particularly true of the automated aviation environments, as they possess many, as yet unexplored elements, and they hold the key to the designs of the future.

Specific Cognitive Processes Underlying Human Decision-Making in Complex Systems

Human Monitoring Behavior

Although the induction of automation into man-machine systems over the past several decades has been primarily technologically based, more recently, investigators within the human factors field have focused on the effects on both workload and performance of removing the human operator more and more from the control loop and placing him/her into the role of passive monitor, for which he/she is ill-suited (Chambers and Nagel, 1985).

Monitoring behavior involves vigilance, also known as sustained attention, "the ability of observers to maintain their focus of attention and to remain alert to stimuli for prolonged periods of time" (Warm, 1990, p. 2). Research within this area is extremely prolific, particularly when one considers its relatively recent history.

Studies in vigilance date back to World War II, when it was noticed that British radar operators missed more and more targets as their time on watch progressed. Norman Mackworth was the first to begin to study this phenomenon systematically, discovering that the radar observers' detection of targets began to decline after approximately thirty minutes on duty (Parasuraman, Warm, and Dember, 1987; Warm, 1990).

Employing a carefully planned experiment, Mackworth (1948) simulated the task of radar monitoring by developing what came to be known as the "Mackworth Clock." As the name implies, the task involved the monitoring of a pointer which moved around the blank face of a clock for a signal involving a skip, or a larger than normal movement. He discovered the vigilance decrement, or the decline over time of the percentage of correct detections of signals. Subsequent research indicated that various manipulations could produce sharp decrements much sooner than the 30-35 minutes first indicated (Parasuraman et al., 1987).

Over the years, many explanations for the vigilance decrement have been postulated, and many investigations have produced seemingly conflicting results, all in all indicating the complexity of human behavior within this domain.

In an attempt to organize this multi-faceted investigative area so as to improve interpretations of various results, Parasuraman (1976) suggested a task taxonomy based upon:

> well-defined information-processing transactions involved in performing vigilance tasks (e.g., access to working memory, updating to expectancies of signal occurrence, etc.), rather than to poorly-defined, broad factors (e.g., boredom, arousal, motivation) common to the monotonous and prolonged aspects of vigilance and other continuous tasks. (Parasuraman et al., 1987, p. 13)

<u>Signal Properties and Monitoring Behavior</u>

Parasuraman and Davies (1977) proposed the categorization of vigilance tasks into four separate dimensions, including <u>signal type</u>, <u>event rate</u>, <u>sensory modality</u>, and <u>complexity</u>. The dimension of signal type was to include successive versus simultaneous (also known as absolute vs. comparative, respectively) presentations of stimuli, as well as separations on a sensory versus cognitive level. Event rate included low, high, or continuous. Sensory modality concentrated upon visual versus auditory, and, finally, source complexity included consideration of single versus multiple signal sources.

Parasuraman, Warm, and Dember (1987) utilized this task classification to organize a paper which attempted to clarify the results of several investigations involving vigilance. Of the various factors examined, several emerged as particularly and consistently important factors in vigilance performance. One such factor is the type of discrimination involved.

The successive versus simultaneous discrimination distinction emerged as a consistently significant factor. With successive, or absolute, presentations of stimuli, signals and non-signals must be compared on the basis of memory, due to the manner of their presentation. As such, they are more subject to limitations of memory. Simultaneous, or comparative, presentations, on the other hand, are not quite as subject to these limitations, as signal and non-signal are presented together. As such, one would expect different results in vigilance performance for tasks which tap into these different processing domains, and indeed, Parasuraman (1976) and Parasuraman and Davies (1977) found highly correlated vigilance performances for different tasks of the same type (successive vs. simultaneous signal discrimination), but not for those of a different type. In considering these, as well as other results, Parasuraman et al. (1987) conclude "that the performances of the same subjects on different vigilance tasks are correlated and that the degree of correlation increases as the information-processing demands that the tasks share in common increase" (p. 16).

Parasuraman et al. (1987) also discuss the distinction between loss of sensitivity and change in response bias in the vigilance decrement, as they differentially affect various tasks.

Sensitivity and response bias or criterion are two elements of Signal Detection Theory (Green and Swets, 1966). The theory outlines two overlapping distributions, signal, and signal plus noise, and describes four separate responses an individual may make to a particular signal: miss, hit, false alarm, and correct rejection. With decrements in the observer's sensitivity, one would expect to see misses and correct rejections rise, while hits and false alarms fell. One would see various results with shifts in the response criterion, or the point at which the observer is willing to decide that a signal is a signal. Depending on whether the response criterion is shifted to be more lenient or more stringent, one would expect rises in false alarms and hits (with subsequent falls in misses and correct rejections) or rises in misses and correct rejections (with subsequent falls in hits and false alarms), respectively.

In examining these elements within the context of successive versus simultaneous task taxonomies, Parasuraman and Davies (1977) and Parasuraman et al. (1987) indicate diminishing sensitivity for successive discrimination tasks with high background event rates, but not for simultaneous discrimination tasks or those with low background event rates. These results were explained "in terms of the greater memory load of the successive-discrimination task, and the sensitivity decrement was attributed to limitations in effortful attention allocation with time" (Parasuraman et al., 1987, p. 16).

Sensitivity is affected, too, by factors which interact with the type of signal discrimination (successive vs. simultaneous). Signal discriminability has a strong effect upon sensitivity, as well as a strong interactive effect when paired with a successive versus a simultaneous task. A sensitivity decrement is more likely to occur with signals which are of low discriminability, which increases the overall resource demand. This effect is even more pronounced when paired with a successive-discrimination task.

Warm (1990), too, discusses the importance of signal discriminability in what he terms signal salience. "A common finding in studies of target detection under alerted conditions is that the frequency of detection is positively related to stimulus amplitude and duration" (p. 8).

In addition to degree of signal discriminability, background event rate, as mentioned earlier, has a consistent effect on sensitivity, and it appears to interact with the type of signal discriminability as well. Background event rate, or "the rate of presentation of stimulus events" (Parasuraman et al., 1987, p. 14). has been shown repeatedly to vary inversely with signal detection. According to Gluckman (1990), "both the speed and accuracy of signal detections vary inversely with event rate, and the vigilance decrement tends to be more pronounced in the context of a fast as compared to a slow event rate" (p. 9). Parasuraman and Davies (1977) describe fast event rate tasks as those involving rates of greater than 24 events per minute. "High-event-rate vigilance tasks have been shown to be resource-consuming" (p. 17), which can explain its adverse interaction with successive-discrimination on sensitivity, as such discrimination involves, again, a greater memory load.

Related to background event rate is the regularity of the event presentation. "Vigilance performance is degraded when events occur irregularly or when signals occur regularly in an irregular event sequence" (Parasuraman et al., 1987, p. 19). This irregularity increases uncertainty as to when (temporal uncertainty) or where (spatial uncertainty) a signal may occur. Such uncertainty increases processing demands, and may decrease both the speed and accuracy with which signals are detected (Parasuraman et al., 1987; Warm, 1990).

Uncertainty, be it spatial or temporal, is influenced by the signal rate, or the frequency of signals presented. The more frequently a signal is presented, the higher its expected probability, the less the uncertainty, and the more likely it is to be detected (Parasuraman et al., 1987; Warm, 1990). Temporal uncertainty, specifically, may be affected by the regularity of the time intervals between signal occurrences, with greater regularity leading to enhanced detectability, in terms of both speed and accuracy. Spatial uncertainty, on the other hand, involves variations in where upon the screen a signal will occur. The more predicable the sequence in which signals appear in various locations, the better the speed and accuracy of detection (Warm, 1990).

Therefore, several factors enter into vigilance performance, and as such each appears to differentially contribute to resource demands. "Thus," according to Parasuraman et al. (1987), "the sensitivity decrement may be determined by the total level of resource demand imposed on the subject" (p. 17).

Cognitive Complexity and Monitoring Behavior

Related to this issue of resource demand is that of cognitive demand, with respect to the complexity of the vigilance task. It is an important area, as monitoring tasks in the real world tend to be more complicated, in terms of both quantity and quality, than those represented in the typical laboratory investigation. This, alone, is a rather difficult domain to examine in that many of the results which have been found appear to be conflicting (Parasuraman et al., 1987; Warm, 1990).

Jerison (1963), for instance, found that the vigilance decrement was aggravated (in that it appeared almost immediately in the watch period) in tasks which required observers to monitor three simultaneous displays. Other investigations, however, found no decrement in tasks employing multiple stimulus sources, but where only one source could appear at one time (Warm, 1990).

In still another investigation, an actual reversal of the vigilance decrement was demonstrated as a result of increased cognitive demand of the vigilance task. Warm, Howe, Fishbein, Dember, and Sprague (1984) define the cognitive demand of a vigilance task in terms of signal complexity, or "the intricacy of the discriminations that observers must make in detecting signals" (p. 15), and it is perhaps this definition (and its applied manipulation) as opposed to one based solely upon number of stimuli or stimulus sources, which contributes to the distinctness of their findings. Their investigation examined the effects of event rate (6 events/min vs. 21 events/min) and the complexity of the signal on vigilance performance. The simple task involved a simple mathematical discrimination, where the complex task involved a complex mathematical discrimination. Whereas performance was fairly stable at low event rates for the simple task, and declined over time at high event rates for the simple tasks, a reversal of the typical vigilance decrement was uncovered for the complex tasks. Vigilance performance actually improved over time for the complex task at both low and high event rates.

Two possible explanations for these results are discussed by Warm et al. (1984): learning and enhanced motivation. In order to test these two hypotheses, a second study was conducted in which two new experimental groups were added to the original two. The first new group (Group 3) was given a list of signals and instructed to respond to any pair presented. The second new group (Group 4) was given a rule which defined critical signals as any pair of digits within a specified range which were typically adjacent in a number sequence, or which were identical. According to the authors, the learning hypothesis would be supported by Groups 3 and 4 beginning at an asymptotic level of performance and remaining somewhat stable over time, as learning beyond an initial level would be unnecessary. On the other hand, if the results could be attributed to enhanced motivation, the authors predicted the reversal of the typical vigilance decrement would occur only in the original Group 2 (complex task), as they speculated that none of the other groups would present enough of an intellectual challenge. Results indicated that all groups demonstrated the typical vigilance decrement over time, except for Group 2, which showed improved performance over time, supporting the hypothesis of enhanced motivation.

Where the previous study would appear to demonstrate the benefits of enhanced intellectual challenge in vigilance performance, a subsequent study conducted by Becker, Gluckman, Warm, and Dember (1987) brings this aspect of cognitive complexity in vigilance tasks into further controversy. These authors found decrements in vigilance performance which they attributed to subjects being exposed to high levels of workload, resulting from mental demand and frustration, for long periods of time. The study employed three experimental groups which varied in terms of the cognitive difficulty of the vigilance task. The first group included a single simultaneous discrimination task, the second group consisted of a dual simultaneous discrimination task, and the third group included a dual successive discrimination task. In this manner, the investigators provided three different tasks of increasing difficulty in terms of capacity demands. Results indicated that sensitivity declined over time for the three groups, and this degradation increased with tasks demanding dual discrimination, particularly in the dual successive discrimination task.

The fundamental conclusion of research within the domain of cognitive complexity within vigilance is simply that whether or not task complexity will enhance, degrade, or stabilize vigilance performance depends upon the circumstances of the investigation. There may even be an inverted-U type of function, in which increases in signal complexity enhance vigilance performance up to a point, after which further increases promote decrements (Parasuraman et al., 1987).

It is obvious from such apparently conflicting results that task complexity, in terms of cognitive demand, probably involves several different factors, with different ones being tapped in different investigations. There may, in fact, be qualitative and quantitative differences in manipulating cognitive complexity on the basis of amount and method of information presentation, the mental processing demands of the task, and the cognitive content of the information itself.

Automatic versus Controlled Processing and Monitoring Behavior

Fisk and Scerbo (1987) have suggested that the many apparent conflicts within the vigilance literature may stem from a lack of understanding of the "mode of information processing" (p. 653) involved in the task.

One influential theory within the realm of cognitive psychology maintains that individuals are capable of processing information by two qualitatively different means: automatic and controlled processing (Schneider and Shiffrin, 1977; Fisk and Scerbo, 1987; Fisk and Schneider, 1981). Much of the work examining the roles of automatic versus control processing in vigilance performance notes the importance of adapting the training and task environments in order to enhance the development of automatic processing and thereby decrease the notorious vigilance decrement. It is speculated that tasks requiring control processing lead to vigilance decrements, whereas those employing automatic processing do not.

The basic differences between automatic and control processing have been highlighted by Fisk and Scerbo (1987):

Automatic Processing:

1. parallel in nature,
2. not limited by STM capacity,
3. requires little or no effort,
4. not under the person's direct control,
5. requires extensive, consistent training to develop,
6. does not modify LTM; and

Control Processing:

1. serial in nature,
2. requires effort,
3. under individual's direct control,
4. requires little or no practice for asymptotic performance,
5. modifies LTM.

Furthermore, Fisk and Scerbo (1987) indicate that the quality, as well as the quantity, of practice differentiates the development of automatic versus control processes. They postulate that stimuli and responses that are consistently mapped, or that require the same overt response to the same stimuli across trials, lead to the development of automatic processing, and improved vigilance performance. Stimuli and responses that possess varied mapping, on the other hand, in which subject responses to the same stimuli change across time, will not lead to automatic processing and will thereby produce a vigilance decrement. This postulation was supported in a later study conducted by Fisk and Schneider (1981).

It would appear, at first glance, that such results conflict with the motivational hypothesis of vigilance performance presented in the afore-mentioned article of Warm et al. (1984). However, Fisk and Schneider (1981) do indicate that it is not control processing alone that tends to create the vigilance decrement, but redundant control processing. This signifies not only the role of mental demand in vigilance performance, but also the role of motivation based upon cognitive challenge, as well as the manner in which the two interact.

Undoubtedly, this expanse of investigative knowledge, when seen as a whole, is not necessarily conflicting, but rather indicates the various factors and levels of performance, as well as the effects of training and the importance of the perspective of a particular situation upon vigilance.

Combined Human-Automation Monitoring Behavior

Recently, in this age of increasing automation, the situation of combined human-computer monitoring has been of interest to several researchers (Corcoran, Dennett, and Carpenter, 1970; 1972; Hollnagel, 1987; Murrell, 1977; Parasuraman, 1987; Weiner and Curry, 1980). Although the potential for automation to reduce the mental demand upon the human operator is recognized, concerns center about the human's ability to recover from automation failures to regain manual control.

Several of the vigilance issues outlined above are addressed by Parasuraman (1987) with respect to the automated environment. Relative to signal frequency and probability, the author suggests that "a potential attentional problem may still remain because the human may be required to detect an infrequent but critical condition" (p.

698). The issue of motivation through cognitive challenge is also addressed, as the author indicates how much of the sensory, preprocessing, and sometimes even higher processing is invaded by artificial sensors and expert systems. The operator is then left to simply monitor the automated system for failures. Even in cases where both the human and computer operators serve monitoring and control functions, Hollnagel (1987) has indicated the need to study the combined effects of such monitoring, as he suggests it is perhaps fundamentally different from human monitoring in a non-automated environment.

There is evidence that the decision making relationship between human and computer monitors in a semi-automated system is complicated by factors such as the attention capacity of the human operator and his/her perceived reliability regarding the computer monitor.

Corcoran, Dennett, and Carpenter (1970) found that, in an investigation of the effects of computer assistance on the recognition of complex sounds, humans performed better when aided and that they relied more upon the aid in conditions of uncertainty (when the sounds were more difficult to hear). Results also indicated a positive relationship between combined monitoring performance and the reliability of the computer aid; however, at extremely high levels of computer aid performance, human performance was reduced (Corcoran, Dennett, and Carpenter, 1972).

Murrell (1977) also found such overreliance in the use of computer aids, with subjects making less use of secondary cue information.

An example of the importance of examining this area of cue utilization when studying aviation is provided by Weiner and Curry's (1980) account of the December 1972 crash of an Eastern Air Lines L-1011:

> During the pre-landing cockpit check, the crew encountered an unsafe landing gear indication (light failed to illuminate). ATC assigned the aircraft to a westward heading at 2000 ft (mean sea level), while the crew attempted to diagnose the problem. The plane was under autopilot control. The flight-crew became preoccupied with the problem at hand (the captain and first officer had pulled the bulb appliance out of the panel to check the lamp, and were having trouble putting it back together). They did not notice that the autopilot had disengaged, and that the aircraft was in a slow descending spiral. They flew into the ground, having never detected their departure from altitude, even with full cockpit instrumentation, extra-cockpit vision, a C-chord altitude alert that sounded (and was present on the cockpit voice recorder), and an ambiguous inquiry from a radar operator in Miami who observed the descent on the alphanumeric read-out on his scope. (p. 1000)

Effects of Automation on Arousal and the Range of Cue Utilization

Investigative interest in the potential narrowing of cue utilization surfaced during the late 1940s and early 1950s. One of the first individuals to spark such interest was D.R. Davis, who in 1948 questioned this variable's relationship to pilot error.

The range of cue utilization is defined by Easterbrook (1958) as "the total number of environmental cues in any situation that an organism observes, maintains an orientation towards, responds to, or associates with a response" (p. 183). He describes this factor as beneficial in certain tasks, with a reduced range of cue utilization improving performance by allowing irrelevant cues to be ignored, thereby optimizing the central task. Easterbrook (1958) also specifies that with certain tasks, such narrowing could be detrimental, if those tasks demand the use of a wider range of cues, and/or if the "drive" to reduce is disorganizing rather than organizing. The "drive" to which Easterbrook (1958) refers concerns, primarily, emotional and/or physiological arousal.

Bahrick, Fitts, and Rankin (1952), whom Easterbrook (1958) credits as the first investigators to examine cue utilization with "deliberate experimental scrutiny" (p.183), observed the phenomena within a context of operant conditioning. The investigation involved a central tracking task, along with peripheral tasks in which subjects were to respond to occasional lights and occasional deflections of a dial (with both lights and dial in the periphery of the central display). It was found that when incentives were offered for heightened performance on the central task, tracking was superior, but performance on the peripheral tasks was diminished, even though potential bonuses were offered for these tasks as well. Thus, there was enhanced concentration upon the central tasks under increased motivation, which improved performance on the central task at the expense of the peripheral tasks.

Bursill (1958) found similar results while manipulating conditions on the basis of a more physiologically based arousal. In his investigation, subjects were to maintain a pointer in a position in which it was superimposed upon another. Peripherally, a semicircular arrangement of six lights provided the secondary task, which required the subjects to detect the illumination of one of the lights. The experiment employed two conditions of ambient temperature: 60-70^0 F versus 95-105^0 F. Results indicated that, under the latter condition, more peripheral signals were missed, whereas performance on the central task became more focused.

Despite observed benefits of such focusing on certain tasks, Easterbrook (1958) indicates that a narrowing of cue utilization does not always have positive results. Specifically, he hypothesizes that any skill which required an "integration of component responses" (p. 187) would suffer under conditions of narrowed cue utilization. "Disintegration of this sort would be expected to ensue from the reduction in range of cue utilization by drive on those tasks in which the receptor-effector span approximated the total range, that is, the most demanding tasks" (p. 187).

Therefore, the tasks that would most likely benefit from a decreased range in cue utilization would be somewhat simple, requiring "relatively little time and information for the adequate preparation of component responses" (Easterbrook, p. 187). In contrast, those tasks which would be expected to suffer would place more demand upon the integration of a variety of informative cues. This would undoubtedly include tasks involving parallel decision making.

Relationship Between Automation, Arousal, Cue-Utilization and Decision-Making

Within the context of task environments that involve a high degree of complexity and potentially dangerous consequences the human operator is subjected to a very demanding task, one in which the arousing nature and the sheer amount of potential information results in a narrowing of cue utilization (Baddeley, 1972; Keinan, 1987; Weltman et al., 1971; Wright, 1974). According to Easterbrook (1958), a task involving such a high degree of complexity should result in decreased performance when the range of cue utilization is reduced.

The scenario of the crash of an L-1011 into the Florida Everglades, described previously, is an extreme but poignant case of over-reliance on an automated component leading to a decreased range in cue utilization, as well as a dramatic change in cue prioritization. The crew's over-reliance on the autopilot set up a situation in which they were made to view the actual task of flying the aircraft as a peripheral task, and the operation of repairing the panel light as a central task, with the characteristic result of focused attention benefitting the latter at the expense of the former. What is uniquely interesting about the aviation environment is that it includes several subtasks involved in a dynamic relationship; one in which respective roles of central and peripheral tasks are not stable but must, in fact, be assigned priorities as part of the overall decision responsibilities of the human operator. This must also be carried out in a potentially time and load stressed environment.

When phenomena such as narrowing of cues and shifting of attention do occur, what determines which cues are to be utilized?

Keinan (1987) suggests that "psychological stress exceeding a certain intensity affects the quality of decision making" (p. 639), indicating that individuals under such conditions may not adhere to "rational-choice models that assume that decisions are based on the weighting of the utilities and probabilities associated with all available courses of action" (p. 639). Instead, the individual under extraordinary stress may, through a process which Janis and Mann (1977) term "hypervigilance," scan available cues in a somewhat erratic fashion, producing incomplete evaluations of the information relevant to the decision at hand.

Keinan (1987) further indicates that there are "three apparently independent ways in which a decision maker's consideration of alternatives might be faulty" (p.639).

The first of these faults is listed as "premature closure", in which "a decision is reached before all available alternatives have been considered" (p. 639). In addition to the research available on narrowing of attention under arousal, there is experimental evidence (Wright, 1974; Wright and Weitz, 1977) which indicates that "the harassed decision maker has trouble assimilating all of the information available to him or her and thus focuses on a limited number of data dimensions" (p. 639).

Keinan (1987) describes the second manner in which a person might be lead to inadequate consideration of the alternatives of a particular decision as a problem with "nonsystematic scanning". Here, she sights evidence that subjects under stress have a scanning process which is disorganized and somewhat erratic (Janis, 1982; Watchel, 1967).

The third possibility that is outlined concerning poor decision making is that of "temporal narrowing", in which "insufficient time is devoted to the consideration of each alternative" (p. 640).

In Keinan's (1987) investigation, subjects performed a multiple choice analogies test under the threat of electric shock. There was no time pressure involved, as she wished to examine the effects of stress alone upon information gathering behaviors. Results indicated a higher incidence of premature closure and nonsystematic scanning under conditions which included threat of electric shock. The investigation did not, however, find support for the author's hypothesis concerning temporal narrowing, but she attributed this to a floor effect.

Under conditions such as are involved in aviation, one would expect there to be a significant level of arousal, particularly in emergency situations. Yet at the same time, one would expect that in order to perform successfully within such an environment the pilot would need to have an organized, systematic method of scanning and, if time did not allow for the consideration of all alternatives, at least some method of prioritizing the relevant information. One might then question whether the existence of such abilities reflects individual differences, learning and experience, or some combination thereof.

Gopher (1982) employed a dichotic listening task to study the possible applications of selective attention tests in predicting success in flight training. He reasoned that flight "involves the ability to focus attention upon demand on relevant aspects of tasks, to switch rapidly from one task to the other, to avoid interference from distracting sources of information, and to divide resources properly in concurrent task performance" (p. 173).

Results indicated that such tests of selective attention could be employed as an independent dimension of overall flight potential predictability. This was in accord with a previous study in which errors in dichotic listening were found to correlate negatively with flight training success (Gopher and Kahneman, 1971).

Such evidence would indicate that there is some element of individual difference in flight capability. However, Gopher (1982) also found that pilots "who had completed a two-year training program had significantly lower error scores on all attention measures" (p. 180), thereby introducing a possible experiential component as well.

Braune and Trollip (1982) have indicated that "although individual differences exist in the sequence in which pilots scan the instruments in a given situation, it appears that pilots in the same situation look for the same categories of information" (p. 996). Additionally, they suggest that pilots do not appear to follow a general, fixed scan pattern. "Experienced instrument pilots adjust their scan patterns to the requirements of a given situation" (p. 996).

Chechile, Eggleston, Fleischman, and Sasseville (1989) suggest that the manner in which pilots gather information from displays reflects "attentional clusters" (p. 35), or the "chunking" of information based upon the "display knowledge network," or "the model of the knowledge representation that results when an experienced or normative viewer first sees the display" (p. 35).

This ability to perceive information in patterns appears to be a general characteristic of expertise, as opposed to a specific attribute of piloting skill.

Chase and Simon (1973), in their examination of novice versus expert chess players, state that "one of the most important processes that underlie chess skill [is] the ability to perceive familiar patterns of pieces" (p. 462).

According to the theory described by Chase and Simon, "a large repertoire of patterns is stored in long-term memory and there is some mechanism that accesses these patterns. In addition, a short-term memory of limited capacity stores the labels (names) of the patterns" (p. 476). Chase and Simon (1973) refer to a "net," quite similar to the frame system described by Minsky (1975), that involves "a set of instructions to the perceptual system for scanning the board systematically for prescribed patterns of pieces" (p. 476). Similar to the matching process of Minsky, a "recognition process continues until attention has been directed to all the salient pieces or short-term memory is filled with labels" (p. 476).

There can be little doubt, then, as to the importance of such patterned attention abilities, governed by the framework of an internal cognitive model, in the case of an aviator making a critical decision.

The Role of Internal Models in Cue Utilization and Decision Making

Braune and Trollip (1982) have speculated that, although the information load upon the pilot can at times be very great, an individual can achieve some level of workload reduction on the basis of an internal model of the situation. Such a model is established through experience and includes "an internal representation of flight phases, which generates certain expectations about given situations" (p. 996). The pilots then construct their scan patterns on the basis of these expectations. This adaptive scanning behavior allows the pilot to compare current information to that expected on the basis of his/her mental model. Therefore, because of the dynamic nature of the environment, the pilot's internal model serves only as a "guide for the information sampling behavior" (p. 997). A mental image of the current situation is constructed on the basis of information gathered through the adaptive scan, and this image is then "compared to the internal model, or expected state of the system, in order to make a decision about control actions" (p. 997). One may begin to understand the nature of such an internal model with an examination of the literature concerning schemata formation.

Schemata Formation

Braune and Trollip (1982) indicate the intricacies involved in the formation of schemata as they possess "variables that may become associated with different aspects of the environment on different occasions" (p. 998), as well as the associative values of those variables. Furthermore, Bruane and Trollip highlight Nelson (1977), who has suggested that such associations, formed on the basis of "recurrent events," facilitate the prediction of "what, when, and why things happen" (p. 997).

Einhorn and Hogarth (1981) discuss a more general relationship between expectancies and the reduction in the number of cues to which one needs to attend. They describe a phenomena termed "cue redundancy" (p. 66), which "reduces the need for attending to and evaluating large numbers of cues" (p. 66). This is said to occur because, "in the natural ecology... cues can indicate the presence of other cues and can thus lead one to expect cue co-occurrences" (p. 66).

An earlier publication by Einhorn, Kleinmuntz and Kleinmuntz (1979) introduced the following potential benefits of cue redundancy:

a. Information search is limited without large losses in predictive accuracy.

b. Attention is highly selective.

c. Dimensionality of the information space is reduced, thereby preventing information overload.

d. Intersubstitutability of cues is facilitated.

e. Unreliability of cues is alleviated by having multiple measures of the same cue variable.

What is very interesting about the development of a pilot's internal model relative to this concept of cue redundancy is that the latter is discussed in relation to the natural ecology. The environment of aviation, on the other hand, is drastically different physiologically, perceptually, and cognitively (Gillingham and Wolfe, 1986) from what humans have evolved and/or learned to expect from the natural ecology. In a sense, pilots must re-evolve, establishing new models; new expectations of a new world so that they can begin to establish a repertoire in cue redundancy, and to continually update and integrate this knowledge into generalized schemata. From such a formulation, they may then direct attention to those dimensions most relevant to the specific situational demands, thereby enhancing the speed and efficiency of their decision making performance.

Investigators examining this relationship between attention and decision making suggest that the allocation of attention plays a vital role in the representation and weighting of various sources of information (Einhorn and Hogarth, 1981; Tversky, 1977).

Einhorn and Hogarth (1981) have described three types of shifts in attention.

These include shifts due to:

1. changes in the point of focus of attention,

2. changes in response mode or framing of the problem,

3. changes in the salience of features due to the specific object set under consideration.

Such selectivity as is noted in the attention process is postulated to occur throughout the decision-making process (Tversky and Kahneman, 1974; Tversky and Kahneman, 1981). Understanding this selectivity is the basic goal of research concerning heuristic principles, the "rules of thumb" in decision making behavior.

Heuristics and Decision-Making

More formally, Merriam-Webster (1986) defines a heuristic as "involving or serving as an aid to learning, discovery, or problem-solving by experimental and esp. trial and error methods" (p. 568). Many of the fundamental principles and biases involved in the use of heuristics have been outlined in the classic work of Tversky and Kahneman (1974). The different principles involved in heuristics are categorized on the basis of how information is selectively accessed, as well as the information processing that is involved.

The representative heuristic involves judgments in which attended information is compared to an abstract model or prototype. A comparison is made between the information being evaluated and the expected features based upon the representative population. Biases associated with the representative heuristic often involve misunderstandings of statistical properties, such as the concepts of randomness and sample size (Tversky and Kahneman, 1974).

The availability heuristic concerns judgments of frequency and likelihood of an event, or the co-occurrence of two or more events. Judgments are based, fundamentally, upon which instances can be most readily brought to the decision-maker's mind. Problems with this heuristic may arise when the most "available" cognitive option is not necessarily the most probable (Tversky and Kahneman, 1974).

Anchoring and adjustment involve estimates that are based upon an initial value and then adjusted until a final judgment is achieved. The final judgments, however, tend bto be influenced by the initial values or endpoints, and biases associated with anchoring and adjustment usually come about because adjustments in these values are often insufficient (Tversky and Kahneman, 1974).

Finally, the simulation heuristic is perhaps best understood as an extension of the availability heuristic into the processing of scripts (Schank and Abelson, 1977). Scripts have been described as "representations of stereotyped events" (Medin and Smith, 1984, p. 132). Whereas the availability heuristic bases its judgments upon the ability of the decision-maker to call upon specific instances, the simulation heuristic extends this concept to include the ease in recalling or even constructing whole scenarios. As is the case with the availability heuristic, biases may arise because the script that is most easily brought to mind may not be the most probable.

Medin and Smith (1984) support the suggestion that a script should perhaps be viewed more as a hierarchy of object concepts than as a single object concept, with a general goal, intermediate scenes, and, finally, actions.

If one equates this view to the aviation environment under investigation, the general goal might be successful flight, the intermediate scenes, the incidents and/or emergencies, and the actions, the control responses. From this point it is not difficult to speculate as to the potential benefits of the various heuristics, as well as their potential downfalls.

The Role of Experience in the Application of Heuristics

Although there are many potential sources of error associated with heuristic biases, they can, if used properly, greatly enhance the efficiency of an otherwise limited information processor. Research suggests that the manner in which heuristics are applied may be related to the level of experience of the decision-maker.

Chase and Simon (1973), for example, have proposed the existence of differences in heuristics use as a function of level of experience. In relation to their work on the development of skill in the game of chess, they suggest that the representative heuristic is more a tool of the expert, whereas the availability heuristic is applied more by novice players. They attribute this to the notion that advanced players have an established repertoire of patterns to be used in a response. With each game, the advanced chess player makes similarity judgments to these patterns or prototypes. The novice, on the other hand, relies on a more limited selection of specific instances which are most readily brought to mind.

In a related line of thinking, the knowledge assembly theory of Hayes-Roth (1977) postulates "that both the representation and processing of knowledge change qualitatively as learning progresses" (p. 260). With experience, the elements of various representations may become linked and strengthened through continuous associations. "A configuration of associated component representations may be strengthened to the point of 'unitization.' Then it is functionally a single element in memory and is activated in an all-or-none fashion as its constituents previously were" (p. 260). One may perhaps even view this as a type of chunking, as efficiency in processing time if accomplished. However, inefficiencies may also be derived if new information is inconsistent with the overall configuration.

The previously discussed role of experience in the use of heuristics can now be examined as to its position within a decision frame, thereby illuminating the possible reasons behind these effects. Minsky (1975) proposes that frames are built and stored in permanent memory as the result of encounters with misfit information.

> I imagine... that a great collection of frame systems is stored in permanent memory, and one of them is evoked when evidence and expectation make it plausible that the scene in view will fit it... I will propose that if a chosen frame does not fit well enough, and if no better one is easily found, and if the matter is important enough, then an adaptation of the best one so far discovered will be constructed and remembered for future use. (p. 214)

Minsky (1975) describes a frame as "a data-structure for representing a stereotyped situation"(p. 212). He suggests these frames are similar to an organized network of levels. The top levels represent constant truths about a particular situation, and therefore vary little. The lower levels, on the other hand, have open slots to be filled as one encounters various specific instances to add to the store of data regarding the situation. One advantage inherent in a frame is its potential ability to reduce the processing demands upon the user. Rather than continually reprocessing our worlds, we develop networks of expected relationships which serve to simplify our analysis of the world as we encounter it. According to Minsky (1975), "a frame's terminals are normally already filled with 'default' assignments" (p. 212) that are derived from our experiences within various environments and situations. Thus, the default assignments described by Minsky "have many uses in representing general information, most likely cases, techniques for bypassing 'logic,' and ways to make useful generalizations" (p. 212). He is careful to point out, however, that the very same default values that can be used to simplify our world by developing stereotypes may at times lead us to err in its analysis. "Properly chosen, such stereotypes could serve as a storehouse of valuable heuristic plan-skeletons; badly selected, they could form paralyzing collections of irrational biases" (pp. 227-228).

Summary

The human decision-maker is constantly sampling the world and comparing the information to expected patterns based upon previously developed frames. The decision-maker tests hypotheses concerning data which cannot be readily determined by the default values of pre-existing frames, and either modifies terminal values or develops altogether new frames in order to accommodate misfit data. In this manner, the frame or internal cognitive model is intricately involved with and affected by the decision-making process that it, in turn, drives and affects. A thorough understanding of how the human decision-maker processes information in any given situation must examine issues related to the framing of the decision process itself. Such issues include the experience level of the decision-maker, the stressors involved in his/her environment, and the stability with which operating frames can be used by the decision-maker to interpret incoming data.

APPENDIX B

SUBJECT BRIEFING PACKAGE

EXPERIMENT I

MEMORANDUM

FROM: LTJG Meghan Carmody Code 6021 X3119

TO: All prospective subjects in Automation and Subtask Decision Making Study.

SUBJ: TASK DESCRIPTION AND SUBJECT BRIEFING

Thank you for agreeing to participate in this study. This briefing package includes some very important information. First, it provides a detailed description of the tasks you will be performing in this experiment. Secondly, it includes your Privacy Act Statement and Consent Form. These are very important and should be read carefully. The Consent Form, in particular highlights the fact that your performance in this study is totally voluntary, and should be a positive experience for you. If you fully understand and agree with the conditions of the Consent Form, then please sign at the bottom of the form on the first day of the experiment.

This experiment is one of a series being performed to assess how different types of cockpit automation effect the performance of various aviation-relevant tasks. The findings will be used to help determine what types of tasks should be automated and how different tasks should be automated.

Please read the briefing very carefully. Although you will be given an initial training session of about an hour, you will find it takes some skill to do all the tasks well. It will help if you have read thoroughly beforehand.

Again, thank you for agreeing to participate in this experiment. I sincerely hope you will find it an interesting and favorable experience. Please feel free to call if you have any questions.

Task Descriptions and Subject Briefing

The tasks which you will be asked to perform are explained below. Also, the final page of the information packet is a diagram of the tasks. You may wish to refer to this while reading.

On the first day of your participation, you will be given a training session of approximately one hour. During this time, the tasks will be verbally described and demonstrated by the experimenter. Please feel free to ask questions at any time during this session. You will also be allowed to practice performing the tasks at this time.

Your participation will total approximately 5 - 6 hours (including training and breaks), but will be spread across two consecutive days. On the first day, you will be given training and three 30 minute experiment sessions. On the second day, you will participate in four 30 minute experiment sessions.

General Instruction

The tasks displayed on the PC are part of a computerized simulation of the kinds of tasks that pilots perform. Each window on the screen represents a different kind of task, as indicated by the headings: System Monitoring, Tracking, and Resource Management.

The sessions in which you will participate include several combinations of automation and manual operation of the three tasks. In each session, either all three tasks will be manually performed, or one of the following tasks will be automated: the System Monitoring or the Resource (Fuel) Management Task.

System Monitoring

All of the information required to perform the monitoring task is displayed in the upper left window of the screen. This task consists of two parts: lights and dials. You will not be monitoring the two lights at the top of this window. You will be monitoring the four dials beneath them for any directional changes in the fluctuation of the pointer. On the first day of your lab run, the experimenter will demonstrate how this task will appear in the "normal" condition.

This task consists of monitoring the four dials. Normally, the yellow pointer within each scale fluctuates within its own specified 3-tick mark range (middle 3 tick marks for the first scale, lower 3 tick marks for the second scale, upper 3 tick marks for the third scale, and middle 3 tick marks for the fourth scale). Your task is to monitor these four dials and detect any change from the normal fluctuation of the pointers. In other words, any time the yellow pointer in any one of the scales goes completely above or completely below its own normal range, this is a signal. You must respond quickly, but accurately, by pressing the appropriate function key: "F1" for the first scale, "F2" for the second scale, "F3" for the third scale, and "F4" for the fourth scale. If you respond correctly to a signal, and do so within 4 seconds, you will receive immediate feedback. The yellow pointer will "jump" back into its own normal range. If you do not respond in time, this will be counted as a "miss". If you respond incorrectly to a signal, or if you respond when no signal is present, this will also be counted against you.

Tracking

The tracking task is displayed in the upper middle section of the screen in the section entitled "Tracking." A joystick will be used to control your position on the screen.

The overall purpose of this task is to keep the airplane, represented by a green circle, within the dotted rectangular area in the center of this task. If you do not control the plane with the joystick, the plane will drift away from the center. Use of the joystick will be demonstrated in your lab training session.

Remember, the overall purpose of the tracking task is to keep the plane in the center rectangular area. Try to maintain this at all times. If the plane leaves the rectangular area, try to return the plane to center as quickly as possible.

Resource (Fuel) Management

The Resource (Fuel) Management task is displayed in the two lower right windows with the headings: "Resource Management" and "Pump Status."

This task is considered a fuel management task. The rectangles are tanks which hold fuel. The shaded green areas within the tanks are fuel, and they increase and decrease with corresponding changes in the fuel level. The small square boxes along the lines which connect the larger fuel tanks are pumps. They transfer fuel from one tank to another in the direction that is indicated by the arrow. The numbers underneath four of the tanks represent the amount of fuel in gallons currently in each of these tanks. They will therefore change as the fuel levels change. The capacity for the two main tanks, A and B, is 4000 gallons each. The supply tanks, C and D, hold a maximum of 2000 gallons each. The supply tanks on the right of each three-tank system have an unlimited capacity -- they never run out.

Your overall goal with this task is to maintain the levels of fuel in tank A at approximately 2200 gallons, and tank B at approximately 2850 gallons. In order to do this, you must transfer fuel to tanks A and B because tanks A and B lose fuel at the rate of 800 gallons per minute. So you can see that without being replenished, these tanks would become empty within minutes. Tanks C and D only lose fuel if they are transferring fuel to another tank.

In your lab training session, the experimenter will demonstrate the process of transferring fuel. Notice that every pump (the small square boxes) has a number and an arrow next to it. The arrow indicates the direction through which fuel can be transferred with that pump. Each pump can only transfer fuel in one direction. The pumps are activated by pressing the key corresponding to the pump that you wish to turn on or off. For example, if you wish to turn on pump number 1 to transfer fuel from tank C to tank A, you simply press the number "1" key across the top of the keyboard.

When you turn the pumps on, two things occur. First, the square on each pump turns green. That means that the pump is actively transferring fuel. When the pump is off, the square is black. The second change on the screen is the numbers that appear in the "Pump Status" window.

Under "Pump Status," two columns of numbers are present. The first column, numbers one through eight, indicate the pump numbers and these correspond directly to the pumps in the diagram. The second column of numbers indicates the flow rates in gallons per minute of each pump when that pump is on. For example, Pump 1 transfers 800 gallons of fuel per minute from tank C to tank A. The flow rate for any given pump is only presented if that pump is on and actively transferring fuel. Pumps 1 and 3 transfer at a rate of 800 gallons per minute, Pumps 7 and 8 transfer at 400 gallons per minute, and the other pumps have flow rates of 600 gallons per minute.

So far, you've seen two conditions for the pumps: on and off. Pressing the pump number key turns the pump on; pressing the key again turns that pump off, and so on. A third condition is the fault condition. You have no control over this. At various times throughout your task, you will see the square on a pump turn red. This means that the pump is inactive as long as that red light is present. You will not be able to turn on this pump until the red light goes out. However, you must be aware that when the fault is corrected and the red light goes out, that pump will automatically be returned to the "off" status (without any light --- black in color). Even if you had turned that pump on before the fault occurred, the pump will not be returned to an "on" condition. You will have to turn it on again if that is what you wish to do.

Your overall goal is to keep the fuel level in Tanks A and B at their respective levels. You may use any strategy that you wish. If the fuel level in these tanks should go outside the desired region, however, please return the fuel level to the target level as soon as possible.

Automated System Monitoring

When the system monitoring task is automated, an "AUTO ON" signal will be displayed in the bottom right of its window. Your task will be one of supervisory monitor, to assure that the automation is operating properly. Occasionally, the automation may function "suboptimally". During your initial lab training session, such a suboptimal function will be demonstrated in order that you may be able to recognize it as such. You may be asked, at the end of a session, to report any such suboptimal functions you may have noticed. There is also a possibility that the automation may fail altogether, at which time no signals (yellow pointer out of bounds) will be corrected during the allotted time period. Such a failure will also be demonstrated to you during your initial lab training session. If such a failure occurs during any of your sessions, you will be required to regain manual operation for the remaining time period.

Automated Resource (Fuel) Management

When the resource management task is automated, the "AUTO ON" signal will be displayed in the bottom right corner of its window. While this task is automated, your task will again be one of supervisory monitor, to ensure that the automation is operating properly. The automation will actively turn on and off pumps in an attempt to maintain the desired fuel levels, just as you would do. Occasionally, there may be suboptimal functioning of the automation. This will be demonstrated in your lab training session. Again, you may be asked, at the end of a session, to report any such suboptimal functions you may have noticed.

There is also a possibility that the automation may fail altogether, at which time no pumps will be activated or deactivated. Such a failure will also be demonstrated to you during your initial training. If such a failure occurs during any of your sessions, you will be required to regain manual operation for the remaining time period.

Workload and Situational Awareness Rating Scales:

The objective here is to capture your perceived "workload" level. The concept of workload is hard to define specifically and is composed of many different aspects. Workload may refer, in part, to the physical demands of the task, the time pressure involved, your expended effort, or your resulting stress or frustration levels.

I would also like to assess your level of "situational awareness." This is also a difficult concept to explicitly define, but it basically refers to your knowledge of changes occurring in the tasks.

You will be asked at the end of each task session to complete two questionnaires. The first assessment will be used to measure your level of situational awareness at different times during the condition. It will be presented via pencil-and-paper. You will read a short statement. Following this, you will indicate the degree to which you agree or disagree with that statement by assigning it a number between 1 and 9 (from decidedly disagree to decidedly agree, respectively). A scale indicating the meanings of all the numbers between 1 and 9 will accompany the question. Please refer to the scale before giving your ratings.

The second assessment will be used to measure workload. Since many factors may be involved, we would like you to tell us about several individual factors rather than one overall workload score. There are six scales on which you will be asked to provide a rating score: mental demand, physical demand, temporal demand, performance, effort, and frustration. The set of six rating scales was developed at the NASA Ames Research Center and has been used in a wide variety of tasks.

Please give your responses thoughtful consideration, but do not spend too much time on them. Your first response will probably accurately reflect your feelings and experiences.

APPENDIX C

SUBJECT BRIEFING PACKAGE

EXPERIMENT II

MEMORANDUM

FROM: LTJG Meghan Carmody Code 6021 X3119

TO: All prospective subjects in Automation and Subtask Decision Making Study.

SUBJ: TASK DESCRIPTION AND SUBJECT BRIEFING

Thank you for agreeing to participate in this study. This briefing package includes some very important information. First, it provides a detailed description of the tasks you will be performing in this experiment. Secondly, it includes your Privacy Act Statement and Consent Form. These are very important and should be read carefully. The Consent Form, in particular highlights the fact that your performance in this study is totally voluntary, and should be a positive experience for you. If you fully understand and agree with the conditions of the Consent Form, then please sign at the bottom of the form on the first day of the experiment.

This experiment is one of a series being performed to assess how different types of cockpit automation effect the performance of various aviation-relevant tasks. The findings will be used to help determine what types of tasks should be automated and how different tasks should be automated.

Please read the briefing very carefully. Although you will be given an initial training session of about an hour, you will find it takes some skill to do all the tasks well. It will help if you have read thoroughly beforehand.

Again, thank you for agreeing to participate in this experiment. I sincerely hope you will find it an interesting and favorable experience. Please feel free to call if you have any questions.

Task Descriptions and Subject Briefing

The tasks which you will be asked to perform are explained below. Also, the final page of the information packet is a diagram of the tasks. You may wish to refer to this while reading.

On the first day of your participation, you will be given a training session of approximately one hour. During this time, the tasks will be verbally described and demonstrated by the experimenter. Please feel free to ask questions at any time during this session. You will also be allowed to practice performing the tasks at this time.

Your participation will total approximately 4 - 5 hours (including training and breaks), but will be spread across two consecutive days. On the first day, you will be given training and three 30 minute experiment sessions. On the second day, you will participate in three additional 30 minute experiment sessions.

General Instruction

The tasks displayed on the PC are part of a computerized simulation of the kinds of tasks that pilots perform. Each window on the screen represents a different kind of task, as indicated by the headings: System Monitoring, Tracking, and Resource Management.

The sessions in which you will participate include several combinations of automation and manual operation of the three tasks. In each session, either all three tasks will be manually performed, or one of the following tasks will be automated: the System Monitoring or the Resource (Fuel) Management Task.

System Monitoring

All of the information required to perform the monitoring task is displayed in the upper left window of the screen. This task consists of two parts: lights and dials. You will not be monitoring the two lights at the top of this window. You will be monitoring the four dials beneath them for any directional changes in the fluctuation of the pointer. On the first day of your lab run, the experimenter will demonstrate how this task will appear in the "normal" condition.

This task consists of monitoring the four dials. Normally, the yellow pointer within each scale fluctuates within its own specified 3-tick mark range (middle 3 tick marks for the first scale, lower 3 tick marks for the second scale, upper 3 tick marks for the third scale, and middle 3 tick marks for the fourth scale). Your task is to monitor these four dials and detect any change from the normal fluctuation of the pointers. In other words, any time the yellow pointer in any one of the scales goes completely above or completely below its own normal range, this is a signal. You must respond quickly, but accurately, by pressing the appropriate function key: "F1" for the first scale, "F2" for the second scale, "F3" for the third scale, and "F4" for the fourth scale. If you respond correctly to a signal, and do so within 4 seconds, you will receive immediate feedback. The yellow pointer will "jump" back into its own normal range. If you do not respond in time, this will be counted as a "miss." If you respond incorrectly to a signal, or if you respond when no signal is present, this will also be counted against you.

Tracking

The tracking task is displayed in the upper middle section of the screen in the section entitled "Tracking." A joystick will be used to control your position on the screen.

The overall purpose of this task is to keep the airplane, represented by a green circle, within the dotted rectangular area in the center of this task. If you do not control the plane with the joystick, the plane will drift away from the center. Use of the joystick will be demonstrated in your lab training session.

Remember, the overall purpose of the tracking task is to keep the plane in the center rectangular area. Try to maintain this at all times. If the plane leaves the rectangular area, try to return the plane to center as quickly as possible.

Resource (Fuel) Management

The Resource (Fuel) Management task is displayed in the two lower right windows with the headings: "Resource Management" and "Pump Status."

This task is considered a fuel management task. The rectangles are tanks which hold fuel. The shaded green areas within the tanks are fuel, and they increase and decrease with corresponding changes in the fuel level. The small square boxes along the lines which connect the larger fuel tanks are pumps. They transfer fuel from one tank to another in the direction that is indicated by the arrow. The numbers underneath four of the tanks represent the amount of fuel in gallons currently in each of these tanks. They will therefore change as the fuel levels change. The capacity for the two main tanks, A and B, is 4000 gallons each. The supply tanks, C and D, hold a maximum of 2000 gallons each. The supply tanks on the right of each three-tank system have an unlimited capacity -- they never run out.

Your overall goal with this task is to maintain the levels of fuel in tank A at approximately 2200 gallons, and tank B at approximately 2850 gallons. In order to do this, you must transfer fuel to tanks A and B because tanks A and B lose fuel at the rate of 800 gallons per minute. So you can see that without being replenished, these tanks would become empty within minutes. Tanks C and D only lose fuel if they are transferring fuel to another tank.

In your lab training session, the experimenter will demonstrate the process of transferring fuel. Notice that every pump (the small square boxes) has a number and an arrow next to it. The arrow indicates the direction through which fuel can be transferred with that pump. Each pump can only transfer fuel in one direction. The pumps are activated by pressing the key corresponding to the pump that you wish to turn on or off. For example, if you wish to turn on pump number 1 to transfer fuel from tank C to tank A, you simply press the number "1" key across the top of the keyboard.

When you turn the pumps on, two things occur. First, the square on each pump turns green. That means that the pump is actively transferring fuel. When the pump is off, the square is black. The second change on the screen is the numbers that appear in the "Pump Status" window.

Under "Pump Status," two columns of numbers are present. The first column, numbers one through eight, indicate the pump numbers and these correspond directly to the pumps in the diagram. The second column of numbers indicates the flow rates in gallons per minute of each pump when that pump is on. For example, Pump 1 transfers 800 gallons of fuel per minute from tank C to tank A. The flow rate for any given pump is only presented if that pump is on and actively transferring fuel. Pumps 1 and 3 transfer at a rate of 800 gallons per minute, Pumps 7 and 8 transfer at 400 gallons per minute, and the other pumps have flow rates of 600 gallons per minute.

So far, you've seen two conditions for the pumps: on and off. Pressing the pump number key turns the pump on; pressing the key again turns that pump off, and so on. A third condition is the fault condition. You have no control over this. At various times throughout your task, you will see the square on a pump turn red. This means that the pump is inactive as long as that red light is present. You will not be able to turn on this pump until the red light goes out. However, you must be aware that when the fault is corrected and the red light goes out, that pump will automatically be returned to the "off" status (without any light --- black in color). Even if you had turned that pump on before the fault occurred, the pump will not be returned to an "on" condition. You will have to turn it on again if that is what you wish to do.

Your overall goal is to keep the fuel level in Tanks A and B at their respective levels. You may use any strategy that you wish. If the fuel level in these tanks should go outside the desired region, however, please return the fuel level to the target level as soon as possible.

Automated System Monitoring

When the system monitoring task is automated, an "AUTO ON" signal will be displayed in the bottom right of its window. Your task will be one of supervisory monitor, to assure that the automation is operating properly. Occasionally, the automation may function "suboptimally." During your initial lab training session, such a suboptimal function will be demonstrated in order that you may be able to recognize it as such. You may be asked, at the end of a session, to report any such suboptimal functions you may have noticed. There is also a possibility that the automation may fail altogether, at which time no signals (yellow pointer out of bounds) will be corrected during the allotted time period. Such a failure will also be demonstrated to you during your initial lab training session. If such a failure occurs during any of your sessions, you will be required to regain manual operation for the remaining time period.

Automated Resource (Fuel) Management

When the resource management task is automated, the "AUTO ON" signal will be displayed in the bottom right corner of its window. While this task is automated, your task will again be one of supervisory monitor, to ensure that the automation is operating properly. The automation will actively turn on and off pumps in an attempt to maintain the desired fuel levels, just as you would do. Occasionally, there may be suboptimal functioning of the automation. This will be demonstrated in your lab training session. Again, you may be asked, at the end of a session, to report any such suboptimal functions you may have noticed.

There is also a possibility that the automation may fail altogether, at which time no pumps will be activated or deactivated. Such a failure will also be demonstrated to you during your initial training. If such a failure occurs during any of your sessions, you will be required to regain manual operation for the remaining time period.

Note: The Tracking Task will never be automated.

Questionnaire Package

At some point in time (between 15 and 30 minutes into the simulation), the simulation will stop, and the blue screen (which will be shown to you in the lab) will appear on the computer screen. At this point in time, you are to immediately begin answering the questionnaire package which will be given to you by the experimenter.

You will be asked about the status and nature of information within the System Monitoring, Resource (Fuel) Management, and Tracking Tasks at the time the simulation stopped. You are to answer the questions as quickly, but as accurately as possible. Do not spend too much time on any one question, and answer all questions, even if you feel you are guessing.

A sample questionnaire will be shown to you in the lab during your training session, and the procedure will be explained verbally. Please feel free to ask any questions you may have at this time.

APPENDIX D

TLX SUBSCALE DEFINITIONS

MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
OWN PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

APPENDIX E

SAGAT QUESTION POOL

System Monitoring

Perception:

(randomly presented)

1. Which scale(s), if any, is(are) presently out of normal range?

   a. No scale is currently out of normal range.

   b. scale 1

   c. scale 2

   d. scale 3

   e. scale 4

2. In the few seconds before the program stopped, the yellow pointer of scale 1 was fluctuating:

   a. in the "lower 3-tick mark" range.

   b. in the "middle 3-tick mark" range.

   c. in the "upper 3-tick mark" range.

3. In the few seconds before the program stopped, the yellow pointer of scale 2 was fluctuating:

   a. in the "lower 3-tick mark" range.

   b. in the "middle 3-tick mark" range.

   c. in the "upper 3-tick mark" range.

4. In the few seconds before the program stopped, the yellow pointer of scale 3 was fluctuating:

   a. in the "lower 3-tick mark" range.

   b. in the "middle 3-tick mark" range.

   c. in the "upper 3-tick mark" range.

5. In the few seconds before the program stopped, the yellow pointer of scale 4 was fluctuating:

   a. in the "lower 3-tick mark" range.

   b. in the "middle 3-tick mark" range.

   c. in the "upper 3-tick mark" range.

Meaning:

1. For the last System Monitoring signal before the program stopped, the appropriate response was:

   a. No response (Not to press any key).

   b. To press F1.

   c. To press F2.

   d. To press F3.

   e. To press F4.

2. If the yellow pointer of scale 1 was in the upper 3 tic-mark area at this moment, I would respond by:

a. Not pressing any key.

b. Pressing the F1 key.

c. Pressing the F2 key.

d. Pressing the F3 key.

e. Pressing the F4 key.

3. If the yellow pointer of scale 2 was in the lower 3 tic-mark area at this moment, I would respond by:

a. Not pressing any key.

b. Pressing the F1 key.

c. Pressing the F2 key.

d. Pressing the F3 key.

e. Pressing the F4 key.

4. If the yellow pointer of scale 3 was in the lower 3 tic-mark area at this moment, I would respond by:

a. Not pressing any key.

b. Pressing the F1 key.

c. Pressing the F2 key.

d. Pressing the F3 key.

e. Pressing the F4 key.

5. If the yellow pointer of scale 4 was in the middle 3 tic-mark area at this moment, I would respond by:

a Not pressing any key.

b. Pressing the F1 key.

c. Pressing the F2 key.

d. Pressing the F3 key.

e. Pressing the F4 key.

6. Within 10 seconds prior to the program stopping, which scale(s), if any, went out of normal range?

a. None.

b. Scale 1.

c. Scale 2.

d. Scale 3.

e. Scale 4.

Future Projection:

1. Approximately how many signals are likely to occur in the next 5 minutes?

a. 0

b. 1-2

c. 3-4

d. 5-6

e. 7-8

Resource Management

Perception:

1. Which pump(s), if any, is(are) currently failed?

   a. None

   b. Pump 1

   c. Pump 2

   d. Pump 3

   e. Pump 4

   f. Pump 5

   g. Pump 6

   h. Pump 7

   i. Pump 8

2. What is the current level of Tank A (approximate if you must)?

3. What is the current level of Tank B (approximate, if you must)?

4. Is Tank A within "normal" limits?

   a. Yes.

   b. No.

5. Is Tank B within "normal" limits?

   a. Yes.

   b. No.

6. What best describes the status of Tank C?

    a. Completely empty.

    b. 0 - 1/4 filled.

    c. 1/4 - 1/2 filled.

    d. 1/2 - 3/4 filled.

    e. Completely filled.

7. What best describes the status of Tank D?

    a. Completely empty.

    b. 0 - 1/4 filled.

    c. 1/4 - 1/2 filled.

    d. 1/2 - 3/4 filled.

    e. Completely filled.

Meaning:

The following questions refer to the status at the time the program stopped:

1. If pump 1 failed at this time, it would be, to your immediate strategy:

    a. a positive occurrence.

    b. a negative occurrence.

    c. of no immediate consequence.

2. If pump 2...

    etc. thru a sample of pumps representing the various areas (supplying the various tanks)

7. With the current situation, what would be the most optimal next step?

8. With the current pump status, the level of fuel in Tank A should:

   a. rise.

   b. fall.

   c. remain the same.

9. With the current pump status, the level of fuel in Tank B should:

   a. rise.

   b. fall.

   c. remain the same.

Future Projection:

1. Approximately how many pump failures are likely to occur in the next 5 minutes?

   a. 0.

   b. 1-2.

   c. 3-4.

   d. 5-6.

   e. 7-8.

APPENDIX F

TRAINING PROTOCOL

Lab Set-Up (Several minutes before subject arrival):

1. Randomize all training, experimental, and control conditions.
2. Turn on MAT task computer (make sure light and white noise on as well).
3. Set up training program.
4. Turn on computer that runs TLX and set-up to TLX screen.
5. Turn on air conditioning if necessary.
6. Unplug telephone.
7. Place "DO NOT ENTER" signs on both laboratory doors.

Subject Set-Up:

1. Subject enters lab.
2. Experimenter introduces self (Experimenter is civil, but professional; minimizes "small talk").
3. Seat subject in chair in front of MAT task computer. Experimenter sits yourself in a chair, facing subject.
4. Ask subject, "Were you able to read the briefing package?" If no, have subject read package and reschedule if necessary. If yes:

5. Inform subject that any questions he/she may have will be answered during the ensuing Training Session: "During training, I will be able to answer any questions you may have. First, I need to get some information from you."

6. Request the briefing package from the subject. Remove the Subject Consent Form; both subject and experimenter sign and date the form.

7. "Now, I'll need to ask you some questions." Read questions aloud from Subject Data Form and fill in appropriate responses.

Subject Training:

1. "Now, we will begin the training sessions. I will first explain all the tasks to you while you observe them on the computer screen. You will then be allowed to practice each task alone for 5 minutes, then the tasks in pairs for 5 minutes, and, finally, all three tasks together for 5 minutes. You will then be given a 5 minute break. Please feel free to ask questions at any time during the training session. "Are you ready to begin?"

2. Once, subject acknowledges that he/she is ready to begin, experimenter press the <RETURN> key to begin the training session. All 3 tasks should appear in manual mode on the screen.

3. Experimenter states, "This is how the tasks will generally appear. As you can see, automation is off in all 3 of these tasks (point to **"AUTO: OFF"** signs in all 3 windows).

4. Experimenter states, "I will explain Manual Operation of the __________ Task first."

5. Example: Resource (Fuel) Management Task has been randomly selected first, followed by System Monitoring and then Tracking:

   "This is the Resource (Fuel) Management Task that you read about in the briefing package. You can see that automation is off in this task (point to **"AUTO: OFF"** sign in the bottom right hand corner of the window). Your goal is to keep the levels of fuel in Tanks A and B within the normal areas (for subjects in Comparative Judgment Type condition, add: which are outlined in red, and point to outlined areas in both tanks). These normal areas cover a range of approximately 2000 to 2400 gallons of fuel for Tank A, and approximately 2650 to 3050 gallons of fuel for Tank B. For optimum performance, you should always maintain the level of fuel as closely as possible to the middle of these normal areas. This means that you always want to stay as close as possible to 2200 gallons of fuel in Tank A, and 2850 gallons of fuel in Tank B. You'll notice that at the bottom of Tanks A and B, the level of fuel currently in the tanks is displayed numerically (point to numbers underneath A and B). In order to maintain the desired level of fuel in A and B, you will need to adopt a strategy of moving fuel among the tanks. This is done by activating and deactivating pumps by pressing the key which corresponds to the number next to the pump you wish to activate or deactivate (demonstrate the activation and deactivation of pumps 1, 2 and 5 and 3, 4 and 6). There are also pumps 7 and 8 (point to 7 and 8) which are only used if it is necessary to transfer fuel between Tanks A and B. Remember that fuel travels only in the direction indicated by the arrow next to each pump (point to all arrows in Systems A and B). Now, whenever a pump is ON, it will be colored green (point to pump(s) that is(are) on. The flow rate of that pump will also be displayed in this window, labeled Pump Status (point to Pump Status window). Pumps 1 and 3 pump the fastest, at a rate of 800 gallons per minute; Pump 2 and 5 and 3 and 6 pump at a rate of 600 gallons per minute; and Pumps 7 and 8 pump at 400 gallons per minute (Demonstrate in Pump Status window). At times, one of the pumps may fail, like this one (point to failed pump). The failure will last 60 seconds, and the pump will be incapable of transferring fuel during this time. The pump will be colored red the entire time it is failed.

"Now, I want to give you a few tips about strategy:

1. You'll notice that when pumps 1 and 3 are on alone, the level of fuel in Tanks A and B, respectively, remains fairly stable. This is fine if the levels are where you want them to be (2200 gallons for Tank A and 2850 gallons for Tanks B). A problem arises if either of these pumps fail, because pumps 2 and 4 cannot pump fuel quickly enough to maintain the levels in Tanks A and B. If such a situation occurs, you may have to revamp your strategy.

2. I don't want to suggest to you to anticipate such failures, and try to maintain an "overload" of fuel in either Tanks A or B, even if at the upper end of their normal ranges. You should always remain as close as possible to 2200 gallons of fuel in Tank A, and 2850 gallons of fuel in Tank B.

6. "Do you have any questions before I explain how the Resource Management Task will be automated?"

7. "When automation is on, the sign in the lower right-hand corner of the task window will read **"AUTO: ON"**. When the task is automated, the system will maintain the levels in Tanks A and B within the normal range. It will probably not remain as close as possible to 2200 and 2850 gallons of fuel as you should, but that is alright. You will be 'locked out' of the pump activation keys, and so you will not be able to turn on or off pumps. There will still be pump failures, and they should last about 60 seconds. If you notice any pump failures lasting longer than 60 seconds, for up to about 90 seconds, this is an instance of "suboptimal automation". Make a mental note of this. You cannot do anything about this, but I may ask you about it later. Also, the levels of fuel in Tanks A and B may go a little above or below normal range at times. Again, you cannot do anything about this, but you may want to make sure automation has not failed completely, which it may or may not do. If automation fails completely, the white **"Auto: On"** sign will turn off, and you will have to regain manual control of the task. Any questions?"

8. "Now, I will allow you to practice manually performing the Resource Management Task alone for 5 minutes."

9. Set up to run 5-minutes of Resource (Fuel) Management Task window alone.

10. "Are you ready?"

    If subject responds, "Yes", press the <RETURN> key to begin the session.

11. Experimenter sits in the designated chair by the unplugged telephone. Remain quiet until the subject has completed the 5-minute session.

12. Ensure that the subject's deviation scores for each tank are no greater than 1000 gallons of fuel for both tanks. If not, explain task again, answer any questions, and rerun. If performance does not meet criteria after second run, cannot use subject.

13. If subject has no questions, then state, "Now, we will move on to the System Monitoring Task."

14. "This is the System Monitoring Task. It is in the Manual Mode. Notice that the sign (point to continuous white sign at the bottom of the window, reading

"**AUTO: OFF**") indicates that automation is off. There are four scales indicating the temperature and pressure of two aircraft engines (point to scales). Within each scale is a yellow pointer. The yellow pointer fluctuates within a normal, 3-tick mark range. This range is different for each scale. The yellow pointer fluctuates normally within the middle 3 tick marks for the first scale (point), the lower 3 tick marks for the second scale (point), the upper 3 tick marks for the third scale (point), and the middle 3 tick marks for the fourth scale (point). So it's middle three, lower three, upper three, middle three. (For subjects in the Comparative Judgment Condition, add: The normal 3-tick mark area for each scale is outlined in red). If at any time any one of the yellow pointers goes completely above or below its own normal range, this is a signal. You must respond to signals as quickly, but accurately as possible, by pressing the corresponding key, which is 'F1' for the first scale (point), 'F2' for the second scale (point), 'F3' for the third scale (point), and 'F4' for the fourth scale (point)."

15. Demonstrate a signal and a response.

16. "As soon as you respond to a signal, the yellow pointer will jump back into normal range. If you do not respond to a signal within 4 seconds, the yellow pointer will return on its own, and this will be recorded as a miss. Also, if you press the wrong key in response to a signal (for example: 'F2' if there is a signal

in scale 1), or if you press a key when there is no signal at all, this will be recorded as a 'false alarm'."

17. "You also want to remember to return your hand in this position at all times when not responding to signals or activating pumps (indicate position)."

18. "Do you have any questions?"

19. "Now I will explain how automation works in the System Monitoring Task."

20. "When the System Monitoring Task is automated, the white, "**Auto: On**" sign will be displayed in the lower, right-hand window. There will still be signals, but the automation should respond to them, bringing them back into normal range within 4 seconds. You will be 'locked out' of the keys and will not be able to respond to signals. Again, the automation should bring any signals back into normal range within 4 seconds. If you notice any signals taking longer than this, up to about 10 seconds, this is an instance of "suboptimal automation". Make a mental note of this. You can't do anything about it, but I may ask you about it later. Also, you might want to make sure automation has not completely failed, which it may or may not do. If automation completely fails, the white, "**Auto: On**" sign will go off, and you will have to begin responding to the signals manually."

21. "Any questions?"

22. If no questions, "Now, I will let you practice this task for 5 minutes."

23. "Are you ready?"

24. If yes, Press <RETURN> to begin. Experimenter sits in the designated chair by unplugged phone until subject has finished. Make sure performance is at least 85% correct (with no more than 1 False Alarm). If not, explain task again, answer any questions, and rerun. If performance is not at least 85% correct on second attempt, cannot use subject.

25. "I will now show you how to operate the Tracking Task.

26. "Remember, you will always be performing the Tracking Task; It will never be automated."

27. If yes, press <RETURN>. Experimenter begins to track while explaining task to subject.

28. "Recall that the goal of the Tracking Task is to keep the green plane centered over the crosshairs in the middle of the rectangular box. To do this, you simply

push the stick left to go left (demonstrate), right to go right (demonstrate), up to go up (demonstrate) and down to go down (demonstrate). When you are within the rectangle, the stick is very responsive. If you let the plane drift out of the rectangle, there is more lag in the response time (demonstrate). Do you have any questions?"

29. If no, "I will now let you practice Tracking alone for 5 minutes".

30. "Are you ready?"

31. If yes, press <RETURN> to begin. Experimenter sits in the designated chair by unplugged phone until subject has finished. Make sure Tracking performance is less than 100 RMSE. If not, explain task again, answer any questions, and rerun. If performance is greater than 100 RMSE on second attempt, do not use subject.

32. "What I will now do is let you practice manual operation of the Resource (Fuel) Management Task with the Tracking Task, then the System Monitoring Task with the Tracking Task, and then all three tasks together. We'll begin with the Resource Management and Tracking Task together".

33. "Are you ready?"

34. If yes, press <RETURN> and sit quietly in the designated chair by the unplugged telephone. When the subject has finished, ensure that deviation for both tanks is no greater than 1600 gallons for the Resource Management Task and that the RMSE for Tracking is no greater than 200.

35. "I will now let you practice the System Monitoring Task with the Tracking Task for five minutes".

36. "Are you ready?"

37. If yes, press <RETURN> and sit quietly in the designated chair by the unplugged telephone. When the subject has finished, ensure performance on System Monitoring is at least 85% (with no more than 1 False Alarm) and RMSE on the Tracking Task is no more than 200.

38. "I will now let you practice all three tasks together for 5 minutes. You will then be given a 5 minute break".

39. "Are you ready?"

40. If yes, press <RETURN> and sit quietly in the designated chair by the unplugged telephone. When the subject has finished, make sure performance on the System

Monitoring Task is at least 75% (with no more than 1 False Alarm), deviation for both tanks on the Resource Management Task is no more than 1600, and RMSE for the Tracking Task is no more than 400.

41. "You may now take break. You may leave the room, but please be back in 5 minutes. You may get something to drink as long as it is not caffeineated."

42. While subject is gone, set up first condition.

APPENDIX G

MEANS, STANDARD ERRORS AND STANDARD DEVIATIONS

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Tracking RMSE Blocks 1-4:

Table 57: Effect(n.s.): Judgment Type (Tracking RMSE Blks 1-4)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	84.44	29.68	7.42
Comparative	79.18	25.23	6.31

Table 58: Effect(sig): Block Number (Tracking RMSE Blks 1-4)

Block Number	Mean RMSE	Standard Error	Standard Deviation
1	71.61	5.44	21.76
2	82.07	6.27	25.10
3	93.98	7.35	29.41
4	85.86	9.21	36.86

Table 59: Effect(n.s.): Condition (Tracking RMSE Blks 1-4)

Condition	Mean RMSE	Standard Error	Standard Deviation
ARMHH	79.75	5.81	23.23
ARMLH	78.83	8.12	32.47
ASMHH	81.45	10.37	41.49
ASMHL	83.83	6.19	24.77
CONHH	86.71	5.99	23.97
CONLH	79.67	7.96	31.86
CONHL	82.44	5.04	20.15

Table 60: Effect(n.s.): Judgment Type x Block (Tracking RMSE Blks 1-4)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	74.39	5.33	21.32
Abs	2	86.43	6.61	26.43
Abs	3	87.03	7.68	30.74
Abs	4	89.90	10.25	41.00
Com	1	61.31	5.46	21.84
Com	2	91.44	6.12	24.07
Com	3	84.25	6.82	27.27
Com	4	79.74	6.94	27.76

Table 61: Effect(n.s.): Judgment Type x Condition (Tracking RMSE Bks 1-4)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	79.09	6.48	25.94
Abs	armlh	78.59	9.30	37.21
Abs	asmhh	86.08	10.08	40.33
Abs	asmhl	85.19	5.72	22.88
Abs	conhh	87.62	5.21	20.82
Abs	conlh	80.38	9.63	38.52
Abs	conhl	84.13	4.90	19.58
Com	armhh	70.41	4.37	17.50
Com	armlh	79.06	5.00	20.01
Com	asmhh	76.82	10.74	42.98
Com	asmhl	82.48	6.74	26.95
Com	conhh	85.80	6.93	27.73
Com	conlh	78.96	5.08	20.33
Com	conhl	80.76	5.29	21.15

Table 62: Effect(sig.): Block x Condition (Tracking RMSE Blks 1-4)

Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
1	armhh	61.38	5.05	20.22
1	armlh	62.66	4.41	17.63
1	asmhh	69.24	8.03	32.11
1	asmhl	74.30	4.56	18.22
1	conhh	62.79	5.47	21.89
1	conlh	67.69	6.09	24.36
1	conhl	76.92	4.47	17.90
2	armhh	82.82	4.74	18.95
2	armlh	74.58	6.10	20.35
2	asmhh	92.62	9.71	35.71
2	asmhl	101.88	6.16	24.66
2	conhh	106.94	6.01	24.03
2	conlh	76.19	6.32	25.28
2	conhl	87.50	5.66	22.63
3	armhh	95.54	7.02	28.07
3	armlh	90.49	8.51	34.05
3	asmhh	72.56	9.37	37.47
3	asmhl	86.21	6.88	27.50
3	conhh	92.37	6.56	26.23
3	conlh	85.74	8.28	33.13
3	conhl	76.59	4.86	19.46

Table 62(Cont.): Effect(sig.): Block x Condition (Tracking RMSE Blks 1-4)

Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
4	armhh	79.25	6.43	25.72
4	armlh	87.58	13.45	53.79
4	asmhh	91.39	15.17	60.69
4	asmhl	72.95	7.18	28.72
4	conhh	84.75	5.94	23.74
4	conlh	89.07	11.17	44.67
4	conhl	88.77	5.17	20.66

Table 63: Effect(sig.): Judgment Type x Block x Condition (Tracking RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
ab	1	armhh	68.63	5.98	23.90
ab	1	armlh	67.62	4.53	18.12
ab	1	asmhh	79.85	5.74	22.94
ab	1	asmhl	85.72	4.28	17.12
ab	1	conhh	63.96	4.66	18.66
ab	1	conlh	70.56	7.31	29.26
ab	1	conhl	84.41	4.78	19.13

Table 63(Cont): Effect(sig.): Judgment Type x Block x Condition (Tracking RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
ab	2	armhh	102.20	4.98	19.93
ab	2	armlh	65.15	7.27	29.07
ab	2	asmhh	84.52	8.20	32.79
ab	2	asmhl	95.25	5.11	20.42
ab	2	conhh	100.54	6.08	24.33
ab	2	conlh	76.64	6.99	27.98
ab	2	conhl	80.69	6.36	25.46
ab	3	armhh	95.84	8.07	32.29
ab	3	armlh	82.47	9.98	39.94
ab	3	asmhh	74.43	14.90	36.50
ab	3	asmhl	87.81	7.31	29.24
ab	3	conhh	96.34	5.39	21.58
ab	3	conlh	89.98	9.72	38.90
ab	3	conhl	82.39	4.18	16.73
ab	4	armhh	89.68	6.91	27.66
ab	4	armlh	99.13	17.94	71.77
ab	4	asmhh	105.54	17.29	69.15
ab	4	asmhl	71.98	6.18	24.73
ab	4	conhh	89.66	4.68	18.71
ab	4	conlh	84.32	14.49	57.98
ab	4	conhl	89.02	4.26	17.05

Table 63(Cont): Effect(sig.): Judgment Type x Block x Condition (Tracking RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
co	1	armhh	54.13	3.40	13.59
co	1	armlh	57.71	4.57	18.28
co	1	asmhh	58.63	9.62	39.48
co	1	asmhl	62.88	5.11	20.44
co	1	conhh	61.62	6.47	25.88
co	1	conlh	64.81	4.61	18.42
co	1	conhl	69.42	4.46	17.84
co	2	armhh	63.45	4.44	17.77
co	2	armlh	84.00	5.09	20.35
co	2	asmhh	100.72	9.71	38.83
co	2	asmhl	108.52	6.98	27.93
co	2	conhh	113.34	6.32	25.30
co	2	conlh	75.73	4.60	18.38
co	2	conhl	94.31	4.96	19.86
co	3	armhh	95.23	4.92	19.70
co	3	armlh	98.52	6.89	27.58
co	3	asmhh	70.69	9.92	39.70
co	3	asmhl	84.61	6.92	27.66
co	3	conhh	88.40	7.95	31.19
co	3	conlh	81.49	5.78	23.12
co	3	conhl	70.79	5.48	21.91

Table 63(Cont.): Effect(sig.): Judgment Type x Block x Condition (Tracking RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
co	4	armhh	68.83	4.71	18.84
co	4	armlh	76.02	3.46	13.84
co	4	asmhh	77.25	13.74	54.94
co	4	asmhl	73.91	7.94	31.76
co	4	conhh	79.84	7.15	28.57
co	4	conlh	93.81	5.35	21.39
co	4	conhl	88.52	6.25	24.98

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Block 2 Tracking RMSE:

Table 64: Effect(n.s.): Judgment Type (Tracking RMSE Blk 2)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	81.50	6.43	25.70
Comparative	76.69	6.01	24.06

Table 65: Effect(sig.): Condition (Tracking RMSE Blk 2)

Condition	Mean RMSE	Standard Error	Standard Deviation
ARMHH	62.66	4.74	18.95
ARMLH	67.69	6.10	24.42
ASMHH	92.62	8.93	35.71
ASMHL	87.50	6.16	24.66
CONHH	86.21	6.01	24.03
CONLH	99.25	6.32	25.28
CONHL	84.75	5.66	22.63

Table 66: Effect(sig.): Judgment Type x Condition (Tracking RMSE Blk 2)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	67.62	4.98	19.93
Abs	armlh	70.56	7.27	29.07
Abs	asmhh	84.52	8.20	32.79
Abs	asmhl	86.69	5.11	20.42
Abs	conhh	87.81	6.08	24.33
Abs	conlh	89.68	6.99	27.98
Abs	conhl	89.65	6.36	25.46

Table 66(Cont.): Effect(sig.): Judgment Type x Condition (Tracking RMSE Blk 2)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Com	armhh	57.71	4.44	17.77
Com	armlh	64.81	5.09	20.35
Com	asmhh	100.72	9.71	38.83
Com	asmhl	94.31	6.98	27.93
Com	conhh	84.61	6.32	25.30
Com	conlh	68.83	4.59	18.38
Com	conhl	79.84	4.96	19.86

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Block 3 Tracking RMSE:

Table 67: Effect(n.s.): Judgment Type (Tracking RMSE Blk 3)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	90.74	7.69	30.75
Comparative	84.29	6.82	27.26

Table 68: Effect(sig.): Condition (Tracking RMSE Blk 3)

Condition	Mean RMSE	Standard Error	Standard Deviation
ARMHH	69.24	7.02	28.07
ARMLH	76.92	8.51	34.47
ASMHH	101.88	9.37	37.41
ASMHL	95.54	6.88	27.50
CONHH	92.37	6.56	26.23
CONLH	87.58	8.28	33.13
CONHL	89.07	4.76	19.46

Table 69: Effect(sig.): Judgment Type x Condition (Tracking RMSE Blk 3)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	79.85	8.07	32.29
Abs	armlh	84.41	9.98	39.94
Abs	asmhh	95.25	9.12	36.50
Abs	asmhl	95.85	7.31	29.24
Abs	conhh	96.34	5.39	21.58
Abs	conlh	99.13	9.72	38.89
Abs	conhl	84.32	4.18	16.73

Table 69(Cont.): Effect(sig.): Judgment Type x Condition (Tracking RMSE Blk 3)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Com	armhh	58.63	4.92	19.70
Com	armlh	69.42	6.89	27.58
Com	asmhh	108.52	9.92	39.70
Com	asmhl	95.23	6.92	27.66
Com	conhh	88.40	7.90	31.19
Com	conlh	76.02	5.78	23.12
Com	conhl	93.81	5.48	21.91

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Block 5 Tracking RMSE:

Table 70: Effect(n.s): Judgment Type (Tracking Blk 5)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	92.11	8.00	32.00
Comparative	81.02	5.48	21.92

Table 71: Effect(n.s.): Condition (Tracking Blk 5)

Condition	Mean RMSE	Standard Error	Standard Deviation
ARMHH	79.85	4.72	18.87
ARMLH	89.35	8.39	33.56
ASMHH	95.93	11.41	45.64
ASMHL	83.59	5.73	22.93
CONHH	82.44	5.21	20.84
CONLH	88.26	8.20	32.75
CONHL	86.55	6.30	25.19

Table 72: Effect(n.s.): Judgment Type x Condition (Tracking RMSE Blk 5)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	82.55	4.64	18.57
Abs	armlh	102.22	10.67	42.67
Abs	asmhh	105.69	14.80	59.19
Abs	asmhl	87.75	6.21	24.87
Abs	conhh	86.56	5.86	23.43
Abs	conlh	98.32	10.45	41.79
Abs	conhl	81.69	3.36	13.46

Table 72(Cont.): Effect(n.s.): Judgment Type x Condition (Tracking RMSE Blk 5)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Com	armhh	77.16	5.01	20.04
Com	armlh	76.47	3.65	14.59
Com	asmhh	86.18	6.81	27.25
Com	asmhl	79.43	5.41	21.65
Com	conhh	78.32	4.63	18.52
Com	conlh	78.21	4.48	17.90
Com	conhl	91.40	8.38	33.54

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Block 6 Tracking RMSE:

Table 73: Effect(n.s.): Judgment Type (Tracking Blk 6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	96.68	12.34	34.92
Comparative	85.91	9.34	26.41

Table 74: Effect(n.s.): Condition (Tracking Blk 6)

Condition	Mean RMSE	Standard Error	Standard Deviation
ARMHH	91.08	6.71	26.82
ARMLH	90.54	11.74	46.95
ASMHH	93.89	7.81	31.24
ASMHL	90.05	7.04	28.18
CONHH	88.47	6.84	27.38
CONLH	86.95	7.26	29.03
CONHL	98.06	7.57	30.29

Table 75: Effect(n.s.): Judgment Type x Condition (Tracking Blk 6)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	94.51	8.15	23.16
Abs	armlh	101.48	22.77	64.40
Abs	asmhh	94.92	12.47	35.26
Abs	asmhl	98.78	11.44	32.37
Abs	conhh	92.16	9.28	26.25
Abs	conlh	95.71	11.92	33.72
Abs	conhl	99.17	10.38	29.37

Table 75(Cont.): Effect(n.s.): Judgment Type x Condition (Tracking Blk 6)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Com	armhh	87.65	11.08	31.35
Com	armlh	79.60	6.14	17.38
Com	asmhh	92.85	10.28	29.07
Com	asmhl	81.32	7.74	21.88
Com	conhh	84.79	10.52	29.77
Com	conlh	78.19	7.85	22.21
Com	conhl	96.95	11.73	33.18

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMHH Tracking RMSE:

Table 76: Effect(n.s.): Judgment Type (Tracking RMSE ARMHH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	79.81	8.57	24.24
Comparative	66.36	7.14	20.20

Table 77: Effect(sig.): Block (Tracking RMSE ARMHH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	61.38	5.06	20.22
2	62.66	4.74	18.95
3	69.24	7.02	28.07
4	74.30	6.43	25.72
5	79.85	4.72	18.87
6	91.08	6.71	26.82

Table 78: Effect(n.s.): Judgment Type x Block (Tracking RMSE ARMHH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	68.63	8.45	23.90
Abs	2	67.62	7.05	19.93
Abs	3	79.85	11.42	32.29
Abs	4	85.72	9.78	27.66
Abs	5	82.55	6.57	18.57
Abs	6	94.51	8.15	23.06
Com	1	54.13	4.81	13.59
Com	2	57.71	6.28	17.77
Com	3	58.63	6.97	19.70
Com	4	62.88	6.68	18.84
Com	5	77.16	7.08	20.04
Com	6	87.65	11.08	31.35

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMLH Tracking RMSE:

Table 79: Effect(n.s.): Judgment Type (Tracking RMSE ARMLH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	87.47	15.67	44.33
Comparative	69.23	6.60	18.67

Table 80: Effect(sig.): Block (Tracking RMSE ARMLH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	62.79	4.41	17.63
2	67.69	6.11	24.42
3	76.92	8.51	34.05
4	82.82	13.45	53.79
5	89.35	8.39	33.56
6	90.54	11.74	46.95

Table 81: Effect(n.s.): Judgment Type x Block (Tracking RMSE ARMLH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	63.96	6.41	18.12
Abs	2	70.56	10.28	29.07
Abs	3	84.41	14.12	39.94
Abs	4	102.20	25.37	71.77
Abs	5	102.22	15.09	42.67
Abs	6	101.48	22.77	64.40
Com	1	61.62	6.46	18.28
Com	2	64.81	7.19	20.35
Com	3	69.42	9.04	27.58
Com	4	63.45	4.88	13.84
Com	5	76.47	5.16	14.59
Com	6	79.60	6.14	17.38

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ASMHH Tracking RMSE:

Table 82: Effect(n.s.): Judgment Type (Tracking RMSE ASMHH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	91.01	15.07	42.63
Comparative	97.60	13.45	38.04

Table 83: Effect(sig.): Block (Tracking RMSE ASMHH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	74.58	8.03	32.11
2	92.62	8.93	35.71
3	101.88	9.37	37.47
4	106.94	15.17	60.69
5	95.93	11.41	45.64
6	93.89	7.81	31.24

Table 84: Effect(n.s.): Judgment Type x Block (Tracking RMSE ASMHH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	65.15	8.11	22.94
Abs	2	84.52	11.59	32.79
Abs	3	95.25	12.90	36.50
Abs	4	100.54	24.45	69.15
Abs	5	105.69	20.93	59.19
Abs	6	94.92	12.47	35.26
Com	1	84.00	13.60	38.48
Com	2	100.72	13.73	38.83
Com	3	108.52	12.97	39.70
Com	4	113.34	19.42	54.94
Com	5	86.18	9.63	27.25
Com	6	92.85	10.28	29.07

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ASMHL Tracking RMSE:

Table 85: Effect(n.s.): Judgment Type (Tracking RMSE ASMHL Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	87.03	8.76	24.79
Comparative	87.42	8.92	25.22

Table 86: Effect(sig.): Block (Tracking RMSE ASMHL Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	76.19	4.56	18.22
2	87.50	6.15	24.66
3	95.54	6.88	27.50
4	90.50	7.18	28.72
5	83.59	5.73	22.93
6	90.05	7.04	28.17

Table 87: Effect(sig.): Judgment Type x Block (Tracking RMSE ASMHL Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	76.64	6.05	17.12
Abs	2	80.69	7.22	20.42
Abs	3	95.84	10.34	29.24
Abs	4	82.47	8.74	24.73
Abs	5	87.75	8.79	24.87
Abs	6	98.78	11.44	32.37
Com	1	75.73	7.23	20.44
Com	2	94.31	9.87	27.66
Com	3	95.23	9.78	27.66
Com	4	98.52	11.23	31.76
Com	5	79.43	7.65	21.65
Com	6	81.32	7.74	21.88

Means, Standard Errors, and Standard Deviations for Effects in
Analysis on Blocks 1-6 CONHH Tracking RMSE:

Table 88: Effect(n.s.): Judgment Type (Tracking RMSE CONHH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	87.88	7.83	22.16
Comparative	81.38	9.40	26.58

Table 89: Effect(sig.): Block (Tracking RMSE CONHH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	72.56	5.47	21.89
2	86.21	7.26	29.03
3	92.37	6.56	26.23
4	85.74	5.94	23.74
5	82.44	5.21	20.84
6	88.47	6.84	27.38

Table 90: Effect(n.s.): Judgment Type x Block (Tracking RMSE CONHH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	74.43	6.60	18.66
Abs	2	87.81	8.60	24.33
Abs	3	96.34	7.63	21.58
Abs	4	89.99	6.62	18.71
Abs	5	86.56	8.28	23.43
Abs	6	92.16	9.28	26.25
Com	1	70.69	9.15	25.88
Com	2	84.61	8.94	25.30
Com	3	88.40	11.03	31.19
Com	4	81.50	10.10	28.57
Com	5	78.32	6.66	18.84
Com	6	84.79	2.83	29.71

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONLH Tracking RMSE:

Table 91: Effect(n.s.): Judgment Type (Tracking RMSE CONLH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	95.13	13.53	38.27
Comparative	74.88	7.16	20.24

Table 92: Effect(sig.): Block (Tracking RMSE CONLH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	76.59	6.09	24.36
2	79.25	6.32	25.28
3	87.58	8.28	33.13
4	91.39	11.17	44.67
5	88.26	8.20	32.75
6	86.95	7.26	29.03

Table 93: Effect(n.s.): Judgment Type x Block (Tracking RMSE CONLH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	82.39	10.34	29.26
Abs	2	89.68	9.89	27.98
Abs	3	99.13	13.75	38.89
Abs	4	105.54	20.50	57.98
Abs	5	98.32	14.78	41.79
Abs	6	95.71	11.92	33.72
Com	1	70.79	6.51	18.42
Com	2	68.83	6.50	18.38
Com	3	76.02	8.17	23.12
Com	4	77.25	7.56	21.39
Com	5	78.21	6.33	17.90
Com	6	78.19	7.85	22.21

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHL Tracking RMSE:

Table 94: Effect(n.s.): Judgment Type (Tracking RMSE CONHL Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	85.97	7.14	20.20
Comparative	87.40	8.92	25.22

Table 95: Effect(sig.): Block (Tracking RMSE CONHL Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	72.95	4.47	17.90
2	84.75	5.65	22.63
3	89.07	4.87	19.46
4	88.77	5.17	20.66
5	86.55	6.30	25.19
6	98.06	7.57	30.30

Table 96: Effect(n.s.): Judgment Type x Block (Tracking RMSE CONHL Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	71.98	6.76	19.13
Abs	2	89.66	9.00	25.46
Abs	3	84.32	5.91	16.73
Abs	4	89.02	6.03	17.05
Abs	5	81.70	4.76	13.46
Abs	6	99.17	10.38	29.37
Com	1	73.91	6.33	17.84
Com	2	79.84	7.02	19.86
Com	3	93.81	7.75	21.91
Com	4	88.52	8.83	24.98
Com	5	91.40	11.86	33.54
Com	6	96.95	11.73	33.18

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Resource (Fuel) Management RMSE Blocks 1-4:

Table 97: Effect(n.s.): Judgment Type (Res Mgmt RMSE Blks 1-4)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	102.65	14.68	41.53
Comparative	98.06	19.51	55.20

Table 98: Effect(sig.): Block (Res Mgmt RMSE Blks 1-4)

Block	Mean RMSE	Standard Error	Standard Deviation
1	80.86	10.33	41.32
2	88.86	14.78	59.10
3	140.17	12.19	48.77
4	91.52	12.96	51.82

Table 99: Effect(sig.): Condition (Res Mgmt RMSE Blks 1-4)

Condition	Mean RMSE	Standard Error	Standard Deviation
asmhh	69.65	14.87	59.49
asmhl	85.47	12.43	49.73
conhh	97.91	12.60	50.38
conlh	141.66	12.92	51.68
conhl	107.08	16.25	65.14

Table 100: Effect(n.s.): Judgment Type x Block (Res Mgmt RMSE Blks 1-4)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	80.92	10.53	29.79
Abs	2	87.50	17.17	48.56
Abs	3	151.30	17.47	49.41
Abs	4	90.88	13.57	38.37
Com	1	80.81	18.12	51.24
Com	2	90.22	22.35	63.22
Com	3	129.04	16.08	45.49
Com	4	92.15	21.51	60.84

Table 101: Effect(n.s.): Judgment Type x Condition (Res Mgmt RMSE Blks 1-4)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	asmhh	66.20	15.41	43.59
Abs	asmhl	82.08	13.88	39.27
Abs	conhh	102.85	10.87	30.74
Abs	conlh	153.25	17.14	48.49
Abs	conhl	108.87	16.10	45.55
Com	asmhh	73.10	16.14	45.64
Com	asmhl	88.85	18.97	53.65
Com	conhh	92.96	22.01	62.26
Com	conlh	130.07	15.19	42.97
Com	conhl	105.29	25.27	71.48

Table 102: Effect(sig.): Block x Condition (Res Mgmt RMSE Blks 1-4)

Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
1	asmhh	57.30	7.95	31.80
1	asmhl	89.42	11.00	44.00
1	conhh	109.28	12.95	51.82
1	conlh	84.63	8.65	34.62
1	conhl	63.69	11.09	44.37
2	asmhh	77.23	9.95	39.79
2	asmlh	100.48	10.24	40.96
2	conhh	82.62	13.97	55.88
2	conlh	62.96	16.81	67.22
2	conhl	121.02	22.92	91.66
3	asmhh	86.03	14.57	58.29
3	asmhl	79.43	9.37	37.49
3	conhh	70.87	8.76	35.03
3	conlh	307.13	14.28	57.11
3	conhl	157.38	14.00	55.98
4	asmhh	58.05	12.96	51.83
4	asmhl	72.54	16.08	64.32
4	conhh	128.85	12.20	48.79
4	conlh	111.92	12.59	50.37
4	conhl	86.23	10.95	43.80

Table 103: Effect(n.s.): Judgment Type x Block x Condition (Res Mgmt RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
ab	1	asmhh	48.71	9.15	25.87
ab	1	asmhl	82.12	10.15	28.70
ab	1	conhh	119.38	13.52	38.23
ab	1	conlh	98.67	10.42	29.48
ab	1	conhl	55.71	9.39	26.57
ab	2	asmhh	76.68	8.18	23.15
ab	2	asmhl	103.08	8.54	24.15
ab	2	conhh	89.38	12.83	36.29
ab	2	conlh	57.05	28.48	80.55
ab	2	conhl	111.33	27.82	78.68
ab	3	asmhh	84.41	24.65	69.72
ab	3	asmhl	80.44	10.91	30.86
ab	3	conhh	68.69	8.59	24.30
ab	3	conlh	339.35	23.68	66.98
ab	3	conhl	183.60	19.50	55.15
ab	4	asmhh	55.01	19.67	55.63
ab	4	asmhl	62.71	25.94	73.36
ab	4	conhh	133.94	8.53	24.14
ab	4	conlh	117.94	6.00	16.96
ab	4	conhl	84.83	7.71	21.80

Table 103(Cont.): Effect(n.s.): Judgment Type x Block x Condition (Res Mgmt RMSE Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
co	1	asmhh	65.88	12.89	36.46
co	1	asmhl	96.71	19.93	56.38
co	1	conhh	99.19	22.95	64.91
co	1	conlh	70.59	14.52	41.08
co	1	conhl	71.68	20.28	57.36
co	2	asmhh	77.78	18.49	52.30
co	2	asmhl	97.89	19.40	54.87
co	2	conhh	75.85	25.40	71.83
co	2	conlh	68.87	10.14	28.67
co	2	conhl	130.72	38.33	108.42
co	3	asmhh	87.66	16.53	46.76
co	3	asmhl	78.43	15.99	45.22
co	3	conhh	73.05	15.94	45.09
co	3	conlh	274.91	10.79	30.51
co	3	conhl	131.16	21.19	59.93
co	4	asmhh	61.10	16.62	47.02
co	4	asmhl	82.38	20.56	58.14
co	4	conhh	123.77	23.76	67.21
co	4	conlh	105.90	25.32	71.61
co	4	conhl	87.62	21.31	60.26

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Resource (Fuel) Management RMSE Block 5:

Table 104: Effect (n.s.): Judgment Type (Res Mgmt RMSE Blk 5)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	130.72	37.71	106.64
Comparative	95.50	21.91	61.97

Table 105: Effect (sig.): Condition (Res Mgmt RMSE Blk 5)

Condition	Mean RMSE	Standard Error	Standard Deviation
armhh	179.46	52.32	209.30
armlh	125.94	17.72	70.90
asmhh	107.96	26.28	105.14
asmhl	79.91	15.34	61.36
conhh	104.69	14.27	57.07
conlh	71.26	16.37	46.30
conhl	122.54	35.15	99.43

Table 106: Effect (n.s.): Judgment Type x Condition (Res Mgmt RMSE Blk 5)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	242.74	100.57	284.46
Abs	armlh	117.21	20.41	57.74
Abs	asmhh	140.51	48.48	137.12
Abs	asmhl	77.56	19.04	53.84
Abs	conhh	127.81	19.19	54.27
Abs	conlh	65.12	9.44	26.69
Abs	conhl	144.11	46.81	132.40
Com	armhh	116.17	21.78	61.60
Com	armlh	134.66	30.13	85.23
Com	asmhh	75.41	17.55	49.65
Com	asmhl	82.27	25.38	71.81
Com	conhh	81.57	18.74	53.02
Com	conlh	77.41	21.78	61.61
Com	conhl	100.98	17.99	50.89

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Resource (Fuel) Management RMSE Block 6:

Table 107: Effect (n.s.): Judgment Type (Res Mgmt RMSE Blk 6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	90.16	19.30	54.58
Comparative	85.09	22.56	63.81

Table 108: Effect (sig.): Condition (Res Mgmt RMSE Blk 6)

Condition	Mean RMSE	Standard Error	Standard Deviation
armhh	116.28	22.44	89.76
armlh	73.49	17.72	70.90
asmhh	114.58	17.52	70.09
asmhl	73.17	12.57	50.27
conhh	79.21	13.42	53.67
conlh	72.76	13.62	54.46
conhl	83.89	13.71	54.85

Table 109: Effect (n.s.): Judgment Type x Condition (Res Mgmt RMSE Blk 6)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	127.72	39.09	110.58
Abs	armlh	63.21	16.55	46.83
Abs	asmhh	141.60	25.98	73.48
Abs	asmhl	68.51	10.13	28.64
Abs	conhh	76.17	15.24	43.11
Abs	conlh	64.67	8.93	25.25
Abs	conhl	89.26	19.16	54.19
Com	armhh	104.85	24.34	68.84
Com	armlh	83.78	18.66	52.77
Com	asmhh	87.56	20.79	58.80
Com	asmhl	77.83	23.83	67.41
Com	conhh	82.26	23.16	65.52
Com	conlh	80.84	26.39	74.63
Com	conhl	78.51	20.75	58.69

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ASMHH Resource (Fuel) Management RMSE:

Table 110: Effect (n.s.): Judgment Type (Res Mgmt RMSE ASMHH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	105.16	22.68	64.16
Comparative	82.56	17.14	48.49

Table 111: Effect (sig.): Block (Res Mgmt RMSE ASMHH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	57.30	7.95	31.80
2	89.42	9.95	39.79
3	109.28	14.58	58.29
4	84.63	12.95	51.83
5	107.96	26.29	105.14
6	114.58	17.52	70.09

Table 112: Effect (n.s.): Judgment Type x Block (Res Mgmt ASMHH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	48.71	9.16	25.87
Abs	2	82.12	8.18	23.15
Abs	3	119.38	24.64	69.72
Abs	4	98.67	19.66	55.63
Abs	5	140.51	48.48	137.12
Abs	6	141.60	25.98	73.48
Com	1	65.88	12.89	36.46
Com	2	96.71	18.49	52.30
Com	3	99.19	16.55	46.76
Com	4	70.59	16.62	47.03
Com	5	75.41	49.65	17.55
Com	6	87.56	20.79	58.80

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
Blocks 1-6 ASMHL Resource (Fuel) Management RMSE:

Table 113: Effect (n.s.): Judgment Type (Res Mgmt RMSE ASMHL Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	78.48	14.11	39.92
Comparative	80.55	20.85	58.97

Table 114: Effect (sig.): Block (Res Mgmt RMSE ASMHL Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	63.69	11.00	44.00
2	77.23	10.24	40.96
3	100.48	9.37	37.49
4	82.62	16.08	64.32
5	79.91	15.34	61.36
6	73.17	12.57	50.27

Table 115: Effect (n.s.): Judgment Type x Block (Res Mgmt RMSE ASMHL Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	55.71	10.15	28.70
Abs	2	76.68	8.54	24.15
Abs	3	103.08	10.91	30.86
Abs	4	89.38	25.94	73.36
Abs	5	77.56	19.02	53.84
Abs	6	68.51	10.11	28.64
Com	1	71.68	19.93	56.38
Com	2	77.78	19.40	54.87
Com	3	97.89	15.99	45.22
Com	4	75.85	20.56	58.14
Com	5	82.27	25.39	71.81
Com	6	77.83	23.83	67.41

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHH Resource (Fuel) Management RMSE:

Table 116: Effect (n.s.): Judgment Type (Res Mgmt RMSE CONHH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	89.53	12.98	36.72
Comparative	88.25	21.66	61.25

Table 117: Effect (sig.): Block (Res Mgmt RMSE CONHH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	62.96	12.96	51.82
2	121.02	13.97	55.88
3	86.03	8.76	35.03
4	79.43	12.20	48.79
5	104.69	14.27	57.07
6	79.21	13.42	53.67

Table 118: Effect (sig.): Judgment Type x Block (Res Mgmt RMSE CONHH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	57.05	13.52	38.23
Abs	2	111.33	12.83	36.29
Abs	3	84.41	8.59	24.30
Abs	4	80.44	8.53	24.14
Abs	5	127.81	19.19	54.27
Abs	6	76.17	15.24	43.11
Com	1	68.87	22.95	64.91
Com	2	130.72	25.46	71.83
Com	3	87.66	15.94	45.09
Com	4	78.43	23.76	67.21
Com	5	81.57	18.74	53.02
Com	6	82.26	23.16	65.52

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
Blocks 1-6 CONLH Resource (Fuel) Management RMSE:

Table 119: Effect (n.s.): Judgment Type (Res Mgmt RMSE CONLH Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	129.41	14.49	40.99
Comparative	116.41	18.15	51.35

Table 120: Effect (sig.): Block (Res Mgmt RMSE CONLH Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	70.87	8.66	34.62
2	307.13	16.81	67.22
3	157.38	14.28	57.11
4	58.05	12.60	50.37
5	71.27	11.58	46.30
6	72.76	13.62	54.46

Table 121: Effect (sig.): Judgment Type x Block (Res Mgmt RMSE CONLH Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	68.69	10.42	29.48
Abs	2	339.35	28.48	80.55
Abs	3	183.60	23.68	66.98
Abs	4	55.01	6.00	16.96
Abs	5	65.12	9.44	26.69
Abs	6	64.68	8.93	25.25
Com	1	73.05	14.52	41.08
Com	2	274.91	10.14	28.67
Com	3	131.16	10.79	30.51
Com	4	61.10	25.32	71.61
Com	5	77.41	21.78	61.61
Com	6	80.84	26.39	74.63

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHL Resource (Fuel) Management RMSE:

Table 122: Effect (n.s.): Judgment Type (Res Mgmt RMSE CONHL Blks 1-6)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	105.47	21.73	61.47
Comparative	96.53	23.31	65.92

Table 123: Effect (sig.): Block (Res Mgmt RMSE CONHL Blks 1-6)

Block	Mean RMSE	Standard Error	Standard Deviation
1	72.54	11.09	44.37
2	128.85	22.92	91.66
3	111.92	14.00	55.98
4	86.23	10.95	43.80
5	122.54	24.86	99.43
6	83.89	13.71	54.85

Table 124: Effect (sig.): Judgment Type x Block (Res Mgmt RMSE CONHL Blks 1-6)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	62.71	9.39	26.57
Abs	2	133.94	27.82	78.68
Abs	3	117.94	19.50	55.15
Abs	4	84.83	7.71	21.80
Abs	5	144.11	46.81	132.40
Abs	6	89.26	19.16	54.19
Com	1	82.38	20.28	57.36
Com	2	123.77	38.33	108.42
Com	3	105.90	21.18	59.93
Com	4	87.62	21.31	60.26
Com	5	100.98	17.99	50.89
Com	6	78.51	20.75	58.69

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring RT Blks 1-4:

Table 125: Effect (n.s.): Judgment Type (System Monitoring RT Blks 1-4)

Judgment Type	Mean RMSE	Standard Error	Standard Deviation
Absolute	3.05	0.62	1.76
Comparative	2.80	0.47	1.32

Table 126: Effect (n.s.): Block (System Monitoring RT Blks 1-4)

Block	Mean RMSE	Standard Error	Standard Deviation
1	2.72	0.38	1.52
2	2.93	0.36	1.42
3	2.97	0.50	1.99
4	3.07	0.38	1.51

Table 127: Effect (n.s.): Condition (System Monitoring RT Blks 1-4)

Condition	Mean RMSE	Standard Error	Standard Deviation
armhh	2.94	0.32	1.28
armlh	2.98	0.34	1.38
conhh	3.14	0.36	1.46
conlh	2.76	0.43	1.74
conhl	2.79	0.55	2.19

Table 128: Effect (n.s.): Judgment Type x Block (System Monitoring RT Blks 1-4)

JT	Block	Mean RMSE	Standard Error	Standard Deviation
Abs	1	3.03	0.64	1.81
Abs	2	3.03	0.49	1.39
Abs	3	3.17	0.67	1.90
Abs	4	2.96	0.59	1.68
Com	1	2.41	0.43	1.21
Com	2	2.83	0.48	1.36
Com	3	2.77	0.55	1.56
Com	4	3.18	0.41	1.16

Table 129: Effect (n.s.): Judgment Type x Condition (System Monitoring RT Blks 1-4)

JT	Cond	Mean RMSE	Standard Error	Standard Deviation
Abs	armhh	3.33	0.59	1.66
Abs	armlh	3.20	0.53	1.49
Abs	conhh	3.43	0.56	1.57
Abs	conlh	2.81	0.77	2.17
Abs	conhl	2.46	0.68	1.93
Com	armhh	2.55	0.20	0.56
Com	armlh	2.76	0.45	1.26
Com	conhh	2.85	0.48	1.36
Com	conlh	2.72	0.39	1.10
Com	conhl	3.12	0.87	2.47

Table 130: Effect (n.s.): Judgment Type x Block x Condition (System Monitoring RT Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
ab	1	armhh	3.11	0.68	1.91
ab	1	armlh	3.38	0.47	1.32
ab	1	conhh	3.98	0.37	1.05
ab	1	conlh	2.80	0.93	2.61
ab	1	conhl	2.89	0.78	2.21
ab	2	armhh	2.54	0.29	0.83
ab	2	armlh	3.74	0.28	0.78
ab	2	conhh	3.21	0.38	1.08
ab	2	conlh	3.38	0.70	1.97
ab	2	conhl	2.25	0.80	2.27
ab	3	armhh	3.68	0.92	2.60
ab	3	armlh	3.79	0.75	2.13
ab	3	conhh	3.32	0.91	2.58
ab	3	conlh	3.16	0.61	1.73
ab	3	conhl	1.89	0.65	1.85
ab	4	armhh	3.99	0.45	1.26
ab	4	armlh	2.91	0.62	1.76
ab	4	conhh	3.19	0.56	1.58
ab	4	conlh	1.89	0.85	2.40
ab	4	conhl	2.81	0.49	1.39

Table 130(Cont.): Effect (n.s.): Judgment Type x Block x Condition (System Monitoring RT Blks 1-4)

JT	Blk	Cond	Mean RMSE	Standard Error	Standard Deviation
co	1	armhh	2.86	0.26	0.73
co	1	armlh	2.05	0.46	1.30
co	1	conhh	2.15	0.40	1.14
co	1	conlh	1.98	0.38	1.08
co	1	conhl	3.01	0.62	1.77
co	2	armhh	2.24	0.08	0.23
co	2	armlh	3.05	0.27	0.77
co	2	conhh	2.32	0.42	1.18
co	2	conlh	3.01	0.43	1.23
co	2	conhl	3.55	1.19	3.37
co	3	armhh	2.50	0.23	0.65
co	3	armlh	3.06	0.66	1.86
co	3	conhh	2.70	0.44	1.26
co	3	conlh	2.81	0.47	1.32
co	3	conhl	2.78	0.95	2.69
co	4	armhh	2.61	0.22	0.63
co	4	armlh	2.87	0.39	1.11
co	4	conhh	4.22	0.44	1.25
co	4	conlh	3.08	0.28	0.77
co	4	conhl	3.13	0.72	2.04

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring RT Block 5:

Table 131: Effect (n.s.): Judgment Type (System Monitoring RT Blk 5)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.65	0.77	2.17
Comparative	3.13	0.52	1.46

Table 132: Effect (n.s.): Condition (System Monitoring RT Blk 5)

Cond	Median RT	Standard Error	Standard Deviation
armhh	3.12	0.32	1.30
armlh	3.39	0.46	1.86
asmhh	2.12	0.28	1.10
asmhl	4.50	0.80	3.18
conhh	3.52	0.46	1.82
conlh	3.60	0.58	2.34
conhl	3.48	0.47	1.87

Table 133: Effect (n.s.): Judgment Type x Condition (System Monitoring RT Blk 5)

JT	Cond	Median RT	Standard Error	Standard Deviation
Abs	armhh	3.18	0.33	0.92
Abs	armlh	3.32	0.71	2.02
Abs	asmhh	1.95	0.39	1.11
Abs	asmhl	5.09	1.34	3.79
Abs	conhh	3.72	0.70	1.98
Abs	conlh	4.27	1.07	3.02
Abs	conhl	4.02	0.84	2.38
Com	armhh	3.05	0.59	1.66
Com	armlh	3.46	0.64	1.81
Com	asmhh	2.30	0.40	1.14
Com	asmhl	3.92	0.90	2.55
Com	conhh	3.33	0.62	1.75
Com	conlh	2.93	0.44	1.24
Com	conhl	2.94	0.38	1.07

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring RT Block 6:

Table 134: Effect (sig.): Judgment Type (System Monitoring RT Blk 6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.50	0.85	2.42
Comparative	2.60	0.47	1.34

Table 135: Effect (n.s.): Condition (System Monitoring RT Blk 6)

Cond	Median RT	Standard Error	Standard Deviation
armhh	3.29	0.42	1.69
armlh	3.15	0.46	1.84
asmhh	2.60	0.31	1.23
asmhl	3.13	0.67	2.68
conhh	2.91	0.47	1.87
conlh	3.71	0.78	3.12
conhl	2.56	0.38	1.53

Table 136: Effect (n.s.): Judgment Type x Condition (System Monitoring RT Blk 6)

JT	Cond	Median RT	Standard Error	Standard Deviation
Abs	armhh	3.53	0.69	1.94
Abs	armlh	3.75	0.79	2.23
Abs	asmhh	2.91	0.52	1.47
Abs	asmhl	2.85	1.22	3.44
Abs	conhh	3.71	0.76	2.14
Abs	conlh	5.03	1.33	3.77
Abs	conhl	2.72	0.69	1.95
Com	armhh	3.04	0.52	1.48
Com	armlh	2.56	0.43	1.23
Com	asmhh	2.28	0.33	0.92
Com	asmhl	3.41	0.65	1.84
Com	conhh	2.10	0.42	1.20
Com	conlh	2.39	0.57	1.60
Com	conhl	2.41	0.39	1.09

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMHH System Monitoring RT

Table 137: Effect (n.s.): Judgment Type (System Monitoring RT ARMHH Blks 1-6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.16	0.55	1.57
Comparative	2.52	0.32	0.91

Table 138: Effect (n.s.): Block (System Monitoring RT ARMHH Blks 1-6)

Block	Median RT	Standard Error	Standard Deviation
1	2.98	0.35	1.41
2	2.21	0.15	0.62
3	3.07	0.52	2.06
4	2.39	0.26	1.05
5	3.12	0.32	1.30
6	3.29	0.42	1.69

Table 139: Effect (n.s.): Judgment Type x Block (System Monitoring RT ARMHH Blks 1-6)

JT	Block	Median RT	Standard Error	Standard Deviation
Abs	1	3.11	0.68	1.91
Abs	2	2.38	0.29	0.83
Abs	3	3.98	0.92	2.60
Abs	4	2.80	0.44	1.26
Abs	5	3.18	0.32	0.92
Abs	6	3.53	0.69	1.94
Com	1	2.86	0.26	0.73
Com	2	2.05	0.08	0.23
Com	3	2.15	0.23	0.65
Com	4	1.98	0.22	0.63
Com	5	3.05	0.59	1.66
Com	6	3.04	0.52	1.48

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMLH System Monitoring RT

Table 140: Effect (n.s.): Judgment Type (System Monitoring RT ARMLH Blks 1-6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.24	0.60	1.71
Comparative	2.77	0.48	1.34

Table 141: Effect (n.s.): Block (System Monitoring RT ARMLH Blks 1-6)

Block	Median RT	Standard Error	Standard Deviation
1	2.95	0.32	1.27
2	2.39	0.19	0.77
3	3.39	0.49	1.96
4	2.77	0.37	1.50
5	3.39	0.46	1.86
6	3.15	0.46	1.84

Table 142: Effect (n.s.): Judgment Type x Block (System Monitoring RT ARMLH Blks 1-6)

JT	Block	Median RT	Standard Error	Standard Deviation
Abs	1	2.89	0.37	1.05
Abs	2	2.54	0.38	1.08
Abs	3	3.74	0.91	2.58
Abs	4	3.21	0.56	1.58
Abs	5	3.32	0.70	1.98
Abs	6	3.75	0.76	2.14
Com	1	3.01	0.46	1.30
Com	2	2.24	0.27	0.77
Com	3	3.05	0.66	1.86
Com	4	2.32	0.39	1.11
Com	5	3.46	0.64	1.81
Com	6.	2.56	0.43	1.23

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHH System Monitoring RT

Table 143: Effect (n.s.): Judgment Type (System Monitoring RT CONHH Blks 1-6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.42	0.61	1.74
Comparative	2.92	0.46	1.30

Table 144: Effect (n.s.): Block (System Monitoring RT CONHH Blks 1-6)

Block	Median RT	Standard Error	Standard Deviation
1	3.20	0.27	1.07
2	2.90	0.32	1.28
3	3.09	0.51	2.06
4	3.43	0.36	1.43
5	3.53	0.45	1.82
6	2.91	0.47	1.87

Table 145: Effect (n.s.): Judgment Type x Block (System Monitoring RT CONHH Blks 1-6)

JT	Block	Median RT	Standard Error	Standard Deviation
Abs	1	3.38	0.37	1.05
Abs	2	2.25	0.38	1.08
Abs	3	3.68	0.91	2.58
Abs	4	3.79	0.56	1.58
Abs	5	3.72	0.70	1.98
Abs	6	3.71	0.76	2.14
Com	1	3.01	0.40	1.14
Com	2	3.55	0.42	1.18
Com	3	2.50	0.44	1.26
Com	4	3.06	0.44	1.25
Com	5	3.33	0.62	1.75
Com	6	2.10	0.42	1.20

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONLH System Monitoring RT

Table 146: Effect (n.s.): Judgment Type (System Monitoring RT CONLH Blks 1-6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	3.61	0.91	2.57
Comparative	2.70	0.42	1.20

Table 147: Effect (n.s.): Block (System Monitoring RT CONLH Blks 1-6)

Block	Median RT	Standard Error	Standard Deviation
1	3.01	0.49	1.95
2	2.98	0.40	1.59
3	3.34	0.39	1.55
4	3.80	0.46	1.86
5	3.60	0.58	2.34
6	3.71	0.78	3.12

Table 148: Effect (n.s.): Judgment Type x Block (System Monitoring RT CONLH Blks 1-6)

JT	Block	Median RT	Standard Error	Standard Deviation
Abs	1	3.32	0.92	2.61
Abs	2	3.16	0.69	1.97
Abs	3	1.89	0.61	1.73
Abs	4	4.00	0.85	2.40
Abs	5	4.27	1.07	3.02
Abs	6	5.03	1.33	3.77
Com	1	2.70	0.38	1.08
Com	2	2.81	0.43	1.23
Com	3	2.78	0.47	1.32
Com	4	2.61	0.27	0.77
Com	5	2.93	0.44	1.24
Com	6	2.39	0.57	1.60

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHL System Monitoring RT

Table 149: Effect (n.s.): Judgment Type (System Monitoring RT CONHL Blks 1-6)

Judgment Type	Median RT	Standard Error	Standard Deviation
Absolute	2.92	0.71	2.01
Comparative	3.11	0.71	2.01

Table 150: Effect (n.s.): Block (System Monitoring RT CONHL Blks 1-6)

Block	Median RT	Standard Error	Standard Deviation
1	2.91	0.48	1.93
2	3.19	0.71	2.82
3	1.89	0.58	2.31
4	2.81	0.42	1.69
5	4.03	0.47	1.87
6	2.72	0.38	1.53

Table 151: Effect (n.s.): Judgment Type x Block (System Monitoring RT CONHL Blks 1-6)

JT	Block	Median RT	Standard Error	Standard Deviation
Abs	1	2.91	1.93	0.48
Abs	2	3.19	2.82	0.71
Abs	3	1.89	2.31	0.58
Abs	4	2.81	1.69	0.42
Abs	5	4.03	1.87	0.47
Abs	6	2.72	1.53	0.38
Com	1	2.87	1.77	0.62
Com	2	4.22	3.37	1.19
Com	3	3.08	2.69	0.95
Com	4	3.13	2.03	0.72
Com	5	2.94	1.07	0.38
Com	6	2.41	1.09	0.38

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring FA Blks 1-4:

Table 152: Effect (n.s.): Judgment Type (System Monitoring FA Blks 1-4)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	0.88	0.42	1.18
Comparative	0.39	0.28	0.79

Table 153: Effect (sig.): Block (System Monitoring FA Blks 1-4)

Block	Mean FA	Standard Error	Standard Deviation
1	0.82	0.32	1.29
2	0.36	0.19	0.78
3	0.91	0.30	1.20
4	0.42	0.28	1.13

Table 154: Effect (n.s.): Condition (System Monitoring FA Blks 1-4)

Condition	Mean FA	Standard Error	Standard Deviation
armhh	0.89	0.36	1.44
armlh	0.53	0.26	1.04
conhh	0.58	0.29	1.17
conlh	0.56	0.28	1.13
conhl	0.59	0.69	2.78

Table 155: Effect (n.s.): Judgment Type x Block (System Monitoring FA Blks 1-4)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	0.98	0.56	1.59
Abs	2	0.48	0.28	0.79
Abs	3	1.42	0.47	1.33
Abs	4	0.62	0.36	1.02
Com	1	0.68	0.19	0.55
Com	2	0.25	0.23	0.64
Com	3	0.40	0.35	0.98
Com	4	0.22	0.35	0.99

Table 156: Effect (n.s.): Judgment Type x Condition (System Monitoring FA Blks 1-4)

JT	Cond	Mean FA	Standard Error	Standard Deviation
Abs	armhh	1.25	0.36	1.03
Abs	armlh	0.78	0.48	1.36
Abs	conhh	0.75	0.47	1.32
Abs	conlh	0.75	0.48	1.35
Abs	conhl	0.84	0.31	0.87
Com	armhh	0.53	0.62	1.75
Com	armlh	0.28	0.17	0.47
Com	conhh	0.41	0.30	0.84
Com	conlh	0.38	0.21	0.59
Com	conhl	0.34	0.11	0.31

Table 157: Effect (n.s.):Block x Condition (System Monitoring FA Blks 1-4)

Blk	Cond	Mean FA	Standard Error	Standard Deviation
1	armhh	1.00	0.36	1.46
1	armlh	0.50	0.55	2.21
1	conhh	0.81	0.14	0.58
1	conlh	0.88	0.28	1.12
1	conhl	0.94	0.25	1.01
2	armhh	0.25	0.21	0.82
2	armlh	0.44	0.14	0.58
2	conhh	0.25	0.20	0.81
2	conlh	0.25	0.30	1.20
2	conhl	0.62	0.12	0.48
3	armhh	1.44	0.38	1.52
3	armlh	0.88	0.20	0.81
3	conhh	0.94	0.38	1.50
3	conlh	0.69	0.33	1.31
3	conhl	0.62	0.22	0.89
4	armhh	0.88	0.50	2.00
4	armlh	0.31	0.14	0.58
4	conhh	0.31	0.44	1.78
4	conlh	0.44	0.22	0.88
4	conhl	0.19	0.10	0.40

Table 158: Effect (n.s.) Judgment Type x Block x Condition (System Monitoring FA Blks 1-4)

JT	Blk	Cond	Mean FA	Standard Error	Standard Deviation
ab	1	armhh	1.38	0.53	1.51
ab	1	armlh	0.38	1.07	3.02
ab	1	conhh	0.88	0.26	0.74
ab	1	conlh	0.62	0.46	1.31
ab	1	conhl	1.62	0.50	1.41
ab	2	armhh	0.38	0.18	0.52
ab	2	armlh	0.62	0.26	0.74
ab	2	conhh	0.38	0.33	0.92
ab	2	conlh	0.38	0.12	0.35
ab	2	conhl	0.62	0.16	0.46
ab	3	armhh	2.00	0.48	1.36
ab	3	armlh	1.50	0.33	0.92
ab	3	conhh	1.50	0.53	1.51
ab	3	conlh	1.25	0.61	1.73
ab	3	conhl	0.88	0.41	1.16
ab	4	armhh	1.25	0.26	0.74
ab	4	armlh	0.62	0.26	0.74
ab	4	conhh	0.25	0.82	2.33
ab	4	conlh	0.75	0.31	0.89
ab	4	conhl	0.25	0.16	0.46

Table 158(Cont.): Effect (n.s.) Judgment Type x Block x Condition (System Monitoring FA Blks 1-4)

JT	Blk	Cond	Mean FA	Standard Error	Standard Deviation
co	1	armhh	0.62	0.50	1.41
co	1	armlh	0.62	0.16	0.46
co	1	conhh	0.75	0.12	0.35
co	1	conlh	1.12	0.18	0.52
co	1	conhl	0.25	0.00	0.00
co	2	armhh	0.12	0.37	1.06
co	2	armlh	0.25	0.12	0.35
co	2	conhh	0.12	0.32	0.92
co	2	conlh	0.12	0.12	0.35
co	2	conhl	0.62	0.18	0.52
co	3	armhh	0.88	0.62	1.75
co	3	armlh	0.25	0.25	0.71
co	3	conhh	0.38	0.48	1.36
co	3	conlh	0.12	0.26	0.74
co	3	conhl	0.38	0.12	0.35
co	4	armhh	0.50	0.99	2.80
co	4	armlh	0.00	0.26	0.74
co	4	conhh	0.38	0.25	0.71
co	4	conlh	0.12	0.27	0.76
co	4	conhl	0.12	0.12	0.35

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring FA Block 5:

Table 159: Effect (sig.): Judgment Type (System Monitoring FA Blk 5)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	0.88	0.42	1.20
Comparative	0.36	0.21	0.60

Table 160: Effect (n.s.): Condition (System Monitoring FA Blk 5)

Condition	Mean FA	Standard Error	Standard Deviation
armhh	0.69	0.20	0.79
armlh	0.75	0.27	1.06
asmhh	1.31	0.59	2.36
asmhl	0.31	0.20	0.79
conhh	0.44	0.20	0.81
conlh	0.31	0.15	0.60
conhl	0.50	0.21	0.82

Table 161: Effect (sig.): Judgment Type x Condition (System Monitoring FA Blk 5)

JT	Cond	Mean FA	Standard Error	Standard Deviation
Abs	armhh	1.00	0.27	0.76
Abs	armlh	1.12	0.48	1.36
Abs	asmhh	2.38	1.07	3.02
Abs	asmhl	0.50	0.38	1.07
Abs	conhh	0.62	0.37	1.06
Abs	conlh	0.25	0.16	0.46
Abs	conhl	0.25	0.25	0.71
Com	armhh	0.38	0.26	0.74
Com	armlh	0.38	0.18	0.52
Com	asmhh	0.25	0.16	0.46
Com	asmhl	0.12	0.12	0.35
Com	conhh	0.25	0.16	0.46
Com	conlh	0.38	0.26	0.74
Com	conhl	0.75	0.31	0.89

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring FA Block 6:

Table 162: Effect (sig.): Judgment Type (System Monitoring FA Blk 6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	1.11	0.44	1.24
Comparative	0.36	0.20	0.57

Table 163: Effect (n.s.): Condition (System Monitoring FA Blk 6)

Condition	Mean FA	Standard Error	Standard Deviation
armhh	0.75	0.21	0.86
armlh	0.75	0.21	0.86
asmhh	0.69	0.31	1.25
asmhl	0.50	0.20	0.82
conhh	0.75	0.46	1.84
conlh	1.31	0.45	1.82
conhl	0.38	0.12	0.50

Table 164: Effect (n.s.): Judgment Type x Condition (System Monitoring FA Blk 6)

JT	Cond	Mean FA	Standard Error	Standard Deviation
Abs	armhh	0.88	0.29	0.83
Abs	armlh	1.00	0.38	1.07
Abs	asmhh	1.38	0.53	1.51
Abs	asmhl	0.88	0.35	0.99
Abs	conhh	1.25	0.90	2.55
Abs	conlh	1.75	0.77	2.19
Abs	conhl	0.62	0.18	0.52
Com	armhh	0.62	0.32	0.92
Com	armlh	0.50	0.19	0.53
Com	asmhh	0.00	0.00	0.00
Com	asmhl	0.12	0.12	0.35
Com	conhh	0.25	0.16	0.46
Com	conlh	0.88	0.48	1.36
Com	conhl	0.12	0.12	0.35

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMHH System Monitoring FA

Table 165: Effect (n.s.): Judgment Type (System Monitoring FA ARMHH Blks 1-6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	0.85	0.34	0.95
Comparative	0.69	0.51	1.44

Table 166: Effect (n.s.): Block (System Monitoring FA ARMHH Blks 1-6)

Block	Mean FA	Standard Error	Standard Deviation
1	1.00	0.36	1.46
2	0.50	0.21	0.82
3	0.81	0.38	1.51
4	0.88	0.50	2.00
5	0.69	0.20	0.79
6	0.75	0.22	0.86

Table 167: Effect (n.s.): Judgment Type x Block (System Monitoring FA ARMHH Blks 1-6)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	1.38	0.53	1.51
Abs	2	0.38	0.18	0.52
Abs	3	0.88	0.48	1.36
Abs	4	0.62	0.26	0.75
Abs	5	1.00	0.27	0.76
Abs	6	0.88	0.29	0.83
Com	1	0.62	0.50	1.41
Com	2	0.62	0.38	1.06
Com	3	0.75	0.62	1.75
Com	4	1.12	0.99	2.80
Com	5	0.38	0.26	0.74
Com	6	0.62	0.32	0.92

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMLH System Monitoring FA

Table 168: Effect (n.s.): Judgment Type (System Monitoring FA ARMLH Blks 1-6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	0.85	0.46	1.30
Comparative	0.27	0.17	0.49

Table 169: Effect (n.s.): Block (System Monitoring FA ARMLH Blks 1-6)

Block	Mean FA	Standard Error	Standard Deviation
1	0.94	0.55	2.21
2	0.25	0.14	0.58
3	0.44	0.20	0.81
4	0.25	0.14	0.58
5	0.75	0.26	1.06
6	0.75	0.22	0.86

Table 170: Effect (n.s.): Judgment Type x Block (System Monitoring FA ARMLH Blks 1-6)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	1.62	1.07	3.02
Abs	2	0.38	0.26	0.74
Abs	3	0.62	0.32	0.92
Abs	4	0.38	0.26	0.74
Abs	5	1.12	0.48	1.36
Abs	6	1.00	0.38	1.07
Com	1	0.25	0.16	0.46
Com	2	0.12	0.12	0.35
Com	3	0.25	0.25	0.71
Com	4	0.12	0.12	0.35
Com	5	0.38	0.18	0.52
Com	6	0.50	0.19	0.53

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHH System Monitoring FA

Table 171: Effect (n.s.): Judgment Type (System Monitoring FA CONHH Blks 1-6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	1.06	0.53	1.49
Comparative	0.40	0.25	0.71

Table 172: Effect (sig.): Block (System Monitoring FA CONHH Blks 1-6)

Block	Mean FA	Standard Error	Standard Deviation
1	0.25	0.14	0.58
2	0.62	0.20	0.81
3	1.44	0.38	1.50
4	0.88	0.44	1.78
5	0.44	0.20	0.81
6	0.75	0.46	1.84

Table 173: Effect (n.s.): Judgment Type x Block (System Monitoring FA CONHH Blks 1-6)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	0.38	0.26	0.74
Abs	2	0.62	0.26	0.74
Abs	3	2.00	0.53	1.51
Abs	4	1.50	0.82	2.33
Abs	5	0.62	0.37	1.06
Abs	6	1.25	0.90	2.55
Com	1	0.12	0.12	0.35
Com	2	0.62	0.32	0.92
Com	3	0.88	0.48	1.36
Com	4	0.25	0.25	0.71
Com	5	0.25	0.16	0.46
Com	6	0.25	0.16	0.46

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONLH System Monitoring FA

Table 174: Effect (n.s.): Judgment Type (System Monitoring FA CONLH Blks 1-6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	1.15	0.47	1.34
Comparative	0.44	0.26	0.74

Table 175: Effect (n.s.): Block (System Monitoring FA CONLH Blks 1-6)

Block	Mean FA	Standard Error	Standard Deviation
1	0.94	0.28	1.12
2	0.69	0.30	1.20
3	0.62	0.33	1.31
4	0.88	0.22	0.88
5	0.31	0.15	0.60
6	1.31	0.46	1.82

Table 176: Effect (n.s.): Judgment Type x Block (System Monitoring FA CONLH Blks 1-6)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	1.50	0.46	1.31
Abs	2	1.25	0.53	1.49
Abs	3	0.88	0.61	1.73
Abs	4	1.25	0.31	0.89
Abs	5	0.25	0.16	0.46
Abs	6	1.75	0.77	2.19
Com	1	0.38	0.18	0.52
Com	2	0.12	0.12	0.35
Com	3	0.38	0.26	0.74
Com	4	0.50	0.27	0.76
Com	5	0.38	0.26	0.74
Com	6	0.88	0.48	1.36

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONLH System Monitoring FA

Table 177: Effect (n.s.): Judgment Type (System Monitoring FA CONLH Blks 1-6)

Judgment Type	Mean FA	Standard Error	Standard Deviation
Absolute	0.46	0.28	0.78
Comparative	0.25	0.14	0.41

Table 178: Effect (n.s.): Block (System Monitoring FA CONLH Blks 1-6)

Block	Mean FA	Standard Error	Standard Deviation
1	0.31	0.25	1.01
2	0.31	0.12	0.48
3	0.44	0.22	0.89
4	0.19	0.10	0.40
5	0.50	0.21	0.82
6	0.38	0.12	0.50

Table 179: Effect (n.s.): Judgment Type x Block (System Monitoring FA CONLH Blks 1-6)

JT	Block	Mean FA	Standard Error	Standard Deviation
Abs	1	0.62	0.50	1.41
Abs	2	0.25	0.16	0.46
Abs	3	0.75	0.41	1.65
Abs	4	0.25	0.16	0.46
Abs	5	0.25	0.25	0.71
Abs	6	0.62	0.18	0.52
Com	1	0.00	0.00	0.00
Com	2	0.38	0.18	0.52
Com	3	0.12	0.12	0.35
Com	4	0.12	0.12	0.35
Com	5	0.75	0.31	0.89
Com	6	0.12	0.12	0.35

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
System Monitoring %Correct Blks 1-4:

Table 180: Effect (sig.): Judgment Type (System Monitoring %Correct Blks 1-4)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.86	0.14	0.41
Comparative	1.25	0.14	0.39

Table 181: Effect (sig.): Block (System Monitoring %Correct Blks 1-4

Block	Mean %Corr	Standard Error	Standard Deviation
1	1.19	0.11	0.44
2	1.20	0.11	0.43
3	0.92	0.12	0.46
4	0.91	0.12	0.47

Table 182: Effect (n.s.): Condition (System Monitoring %Correct Blks 1-4)

Condition	Mean %Corr	Standard Error	Standard Deviation
armhh	1.08	0.14	0.41
armlh	1.14	0.09	0.36
conhh	1.02	0.42	0.11
conlh	1.00	0.13	0.50
conhl	1.03	0.14	0.54

Table 183: Effect (sig.): Judgment Type x Block (System Monitoring %Correct Blks 1-4)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	0.99	0.13	0.37
Abs	2	1.11	0.15	0.42
Abs	3	0.73	0.12	0.34
Abs	4	0.60	0.16	0.44
Com	1	1.40	0.12	0.35
Com	2	1.29	0.13	0.37
Com	3	1.10	0.16	0.44
Com	4	1.22	0.14	0.40

Table 184: Effect (n.s.): Judgment Type x Condition (System Monitoring %Correct Blks 1-4)

JT	Cond	Mean %Corr	Standard Error	Standard Deviation
Abs	armhh	0.88	0.13	0.38
Abs	armlh	0.92	0.13	0.37
Abs	conhh	0.86	0.12	0.33
Abs	conlh	0.77	0.18	0.51
Abs	conhl	0.86	0.16	0.46
Com	armhh	1.28	0.11	0.30
Com	armlh	1.37	0.12	0.34
Com	conhh	1.17	0.16	0.45
Com	conlh	1.24	0.15	0.41
Com	conhl	1.20	0.16	0.44

Table 185: Effect (sig.): Block x Condition (System Monitoring %Correct Blks 1-4)

Blk	Cond	Mean %Corr	Standard Error	Standard Deviation
1	armhh	1.13	0.10	0.38
1	armlh	1.25	0.09	0.35
1	conhh	1.19	0.10	0.41
1	conlh	1.11	0.11	0.44
1	conhl	1.28	0.15	0.60
2	armhh	1.41	0.10	0.39
2	armlh	1.25	0.07	0.30
2	conhh	1.20	0.11	0.46
2	conlh	1.18	0.13	0.51
2	conhl	0.96	0.13	0.51
3	armhh	0.79	0.11	0.42
3	armlh	1.04	0.09	0.34
3	conhh	0.87	0.10	0.41
3	conlh	1.04	0.13	0.54
3	conhl	0.84	0.15	0.59
4	armhh	0.98	0.11	0.45
4	armlh	1.03	0.12	0.47
4	conhh	0.80	0.11	0.43
4	conlh	0.68	0.13	0.53
4	conhl	1.05	0.12	0.50

Table 186: Effect (n.s.): Judgment Type x Block x Condition (System Monitoring %Correct Blks 1-4)

JT	Blk	Cond	Mean %Corr	Standard Error	Standard Deviation
ab	1	armhh	0.90	0.11	0.32
ab	1	armlh	1.01	0.13	0.37
ab	1	conhh	1.06	0.14	0.40
ab	1	conlh	0.82	0.16	0.46
ab	1	conhl	1.14	0.22	0.61
ab	2	armhh	1.39	0.13	0.38
ab	2	armlh	1.10	0.12	0.34
ab	2	conhh	1.16	0.13	0.37
ab	2	conlh	1.11	0.20	0.56
ab	2	conhl	0.79	0.17	0.48
ab	3	armhh	0.55	0.17	0.47
ab	3	armlh	0.87	0.11	0.30
ab	3	conhh	0.67	0.15	0.13
ab	3	conlh	0.89	0.19	0.53
ab	3	conhl	0.66	0.09	0.25
ab	4	armhh	0.65	0.13	0.36
ab	4	armlh	0.71	0.17	0.48
ab	4	conhh	0.55	0.14	0.41
ab	4	conlh	0.26	0.17	0.47
ab	4	conhl	0.84	0.17	0.48

Table 186(Cont.): Effect (n.s.): Judgment Type x Block x Condition (System Monitoring % Blks 1-4)

JT	Blk	Cond	Mean %Corr	Standard Error	Standard Deviation
Co	1	armhh	1.35	0.11	0.30
Co	1	armlh	1.50	0.10	0.27
Co	1	conhh	1.32	0.16	0.44
Co	1	conlh	1.39	0.12	0.35
Co	1	conhl	1.42	0.14	0.39
Co	2	armhh	1.42	0.07	0.21
Co	2	armlh	1.39	0.10	0.27
Co	2	conhh	1.24	0.18	0.50
Co	2	conlh	1.26	0.15	0.43
Co	2	conhl	1.13	0.16	0.44
Co	3	armhh	1.03	0.13	0.36
Co	3	armlh	1.22	0.12	0.34
Co	3	conhh	1.07	0.17	0.47
Co	3	conlh	1.19	0.18	0.52
Co	3	conhl	1.01	0.18	0.52
Co	4	armhh	1.32	0.12	0.34
Co	4	armlh	1.36	0.17	0.49
Co	4	conhh	1.04	0.14	0.40
Co	4	conlh	1.10	0.13	0.36
Co	4	conhl	1.26	0.15	0.44

Means, Standard Errors, and Standard Deviations for Effects in Analysis on System Monitoring %Correct Block 5:

Table 187: Effect (sig.): Judgment Type (System Monitoring %Correct Blk 5)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.73	0.17	0.48
Comparative	1.10	0.15	0.42

Table 188: Effect (sig.): Condition (System Monitoring %Correct Blk 5)

Condition	Mean %Corr	Standard Error	Standard Deviation
armhh	1.06	0.10	0.39
armlh	0.84	0.12	0.47
asmhh	0.73	0.12	0.49
asmhl	0.74	0.14	0.54
conhh	0.99	0.12	0.48
conlh	1.03	0.12	0.46
conhl	1.00	0.14	0.54

Table 189: Effect (n.s.): Judgment Type x Condition (System Monitoring %Correct Blk 5)

JT	Cond	Mean %Corr	Standard Error	Standard Deviation
Abs	armhh	0.88	0.12	0.35
Abs	armlh	0.70	0.21	0.58
Abs	asmhh	0.51	0.14	0.41
Abs	asmhl	0.48	0.18	0.49
Abs	conhh	0.79	0.16	0.44
Abs	conlh	0.87	0.16	0.46
Abs	conhl	0.85	0.22	0.61
Com	armhh	1.25	0.13	0.36
Com	armlh	0.97	0.11	0.30
Com	asmhh	0.94	0.18	0.50
Com	asmhl	1.00	0.16	0.44
Com	conhh	1.20	0.16	0.45
Com	conlh	1.19	0.15	0.43
Com	conhl	1.15	0.16	0.45

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
System Monitoring %Correct Block 6:

Table 190: Effect (sig.): Judgment Type (System Monitoring %Correct Blk 6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.83	0.17	0.48
Comparative	1.18	0.15	0.43

Table 191: Effect (sig.): Condition (System Monitoring %Correct Blk 6)

Condition	Mean %Corr	Standard Error	Standard Deviation
armhh	1.18	0.10	0.41
armlh	1.16	0.11	0.44
asmhh	0.97	0.12	0.46
asmhl	0.74	0.14	0.56
conhh	1.07	0.12	0.48
conlh	0.81	0.12	0.48
conhl	1.11	0.14	0.57

Table 192: Effect (n.s.): Judgment Type x Condition (System Monitoring %Correct Blk 6)

JT	Cond	Mean %Corr	Standard Error	Standard Deviation
Abs	armhh	1.05	0.16	0.45
Abs	armlh	1.08	0.15	0.43
Abs	asmhh	0.80	0.14	0.40
Abs	asmhl	0.44	0.19	0.53
Abs	conhh	0.80	0.12	0.34
Abs	conlh	0.78	0.20	0.57
Abs	conhl	0.86	0.23	0.64

Table 192(Cont.): Effect (n.s.): Judgment Type x Condition (System Monitoring %Correct Blk 6)

JT	Cond	Mean %Corr	Standard Error	Standard Deviation
Com	armhh	1.32	0.13	0.36
Com	armlh	1.23	0.17	0.48
Com	asmhh	1.13	0.17	0.49
Com	asmhl	1.04	0.15	0.43
Com	conhh	1.34	0.16	0.45
Com	conlh	0.83	0.14	0.41
Com	conhl	1.36	0.14	0.39

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMHH System Monitoring %Correct

Table 193: Effect (sig.): Judgment Type (System Monitoring %Correct ARMHH Blks 1-6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.95	0.14	0.39
Comparative	1.35	0.11	0.32

Table 194: Effect (n.s.): Block (System Monitoring %Correct ARMHH Blks 1-6)

Block	Mean %Corr	Standard Error	Standard Deviation
1	1.13	0.09	0.38
2	1.25	0.10	0.39
3	1.19	0.11	0.42
4	1.11	0.11	0.45
5	1.06	0.10	0.39
6	1.18	0.10	0.41

Table 195: Effect (n.s.): Judgment Type x Block (System Monitoring %Correct ARMHH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	0.90	0.11	0.32
Abs	2	1.01	0.13	0.38
Abs	3	1.06	0.16	0.47
Abs	4	0.82	0.13	0.36
Abs	5	0.88	0.12	0.35
Abs	6	1.05	0.16	0.45

Table 195(Cont.): Effect (n.s.): Judgment Type x Block (System Monitoring %Correct ARMHH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Com	1	1.35	0.11	0.30
Com	2	1.50	0.07	0.21
Com	3	1.32	0.13	0.36
Com	4	1.39	0.12	0.34
Com	5	1.25	0.13	0.36
Com	6	1.32	0.13	0.36

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 ARMLH System Monitoring %Correct

Table 196: Effect (n.s.): Judgment Type (System Monitoring %Correct ARMLH Blks 1-6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	1.10	0.15	0.42
Comparative	1.28	0.13	0.36

Table 197: Effect (sig.): Block (System Monitoring %Correct ARMLH Blks 1-6)

Block	Mean %Corr	Standard Error	Standard Deviation
1	1.28	0.09	0.35
2	1.41	0.07	0.30
3	1.25	0.09	0.34
4	1.20	0.12	0.47
5	0.84	0.12	0.47
6	1.16	0.11	0.44

Table 198: Effect (n.s.): Judgment Type x Block (System Monitoring %Correct ARMLH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	1.14	0.13	0.37
Abs	2	1.39	0.12	0.34
Abs	3	1.10	0.11	0.30
Abs	4	1.17	0.17	0.48
Abs	5	0.70	0.21	0.58
Abs	6	1.08	0.15	0.43

Table 198(Cont.): Effect (n.s.): Judgment Type x Block (System Monitoring %Correct ARMLH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Com	1	1.42	0.10	0.27
Com	2	1.42	0.10	0.27
Com	3	1.39	0.12	0.34
Com	4	1.24	0.17	0.49
Com	5	0.97	0.11	0.30
Com	6	1.23	0.17	0.48

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHH System Monitoring %Correct

Table 199: Effect (sig.): Judgment Type (System Monitoring %Correct CONHH Blks 1-6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.82	0.12	0.35
Comparative	1.20	0.16	0.45

Table 200: Effect (sig.): Block (System Monitoring %Correct CONHH Blks 1-6)

Block	Mean %Corr	Standard Error	Standard Deviation
1	1.18	0.10	0.41
2	0.96	0.11	0.46
3	0.79	0.10	0.41
4	1.04	0.11	0.44
5	0.99	0.12	0.48
6	1.07	0.12	0.48

Table 201: Effect (n.s.): Judgment Type x Block (System Monitoring %Correct CONHH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	1.11	0.14	0.40
Abs	2	0.79	0.13	0.37
Abs	3	0.55	0.05	0.13
Abs	4	0.87	0.14	0.41
Abs	5	0.79	0.16	0.44
Abs	6	0.80	0.12	0.34

Table 201(Cont.): Effect (n.s.): Judgment Type x Block (System Monitoring %Correct CONHH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Com	1	1.26	0.16	0.44
Com	2	1.13	0.18	0.50
Com	3	1.03	0.17	0.47
Com	4	1.22	0.14	0.40
Com	5	1.20	0.16	0.45
Com	6	1.34	0.16	0.45

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONLH System Monitoring %Correct

Table 202: Effect (n.s.): Judgment Type (System Monitoring %Correct CONLH Blks 1-6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.76	0.18	0.51
Comparative	1.10	0.15	0.42

Table 203: Effect (n.s.): Block (System Monitoring %Correct CONLH Blks 1-6)

Block	Mean %Corr	Standard Error	Standard Deviation
1	0.87	0.11	0.44
2	1.04	0.13	0.51
3	0.84	0.13	0.53
4	0.98	0.13	0.53
5	1.03	0.12	0.46
6	0.81	0.12	0.48

Table 204: Effect (n.s.): Judgment Type x Block (System Monitoring %Correct CONLH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	0.67	0.16	0.46
Abs	2	0.89	0.20	0.56
Abs	3	0.66	0.19	0.53
Abs	4	0.65	0.17	0.47
Abs	5	0.87	0.16	0.46
Abs	6	0.78	0.20	0.57

Table 204(Cont.): Effect (n.s.): Judgment Type x Block (System Monitoring %Correct CONLH Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Com	1	1.07	0.12	0.34
Com	2	1.19	0.15	0.43
Com	3	1.01	0.18	0.52
Com	4	1.32	0.13	0.36
Com	5	1.19	0.15	0.43
Com	6	0.83	0.14	0.41

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Blocks 1-6 CONHL System Monitoring %Correct

Table 205: Effect (sig.): Judgment Type (System Monitoring %Correct CONHL Blks 1-6)

Judgment Type	Mean %Corr	Standard Error	Standard Deviation
Absolute	0.68	0.18	0.51
Comparative	1.21	0.15	0.43

Table 206: Effect (sig.): Block (System Monitoring %Correct CONHL Blks 1-6)

Block	Mean %Corr	Standard Error	Standard Deviation
1	1.03	0.15	0.60
2	0.80	0.13	0.51
3	0.68	0.15	0.59
4	1.05	0.12	0.49
5	1.00	0.14	0.54
6	1.11	0.14	0.57

Table 207: Effect (n.s.): Judgment Type x Block (System Monitoring %Correct CONHL Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Abs	1	0.71	0.22	0.61
Abs	2	0.55	0.17	0.48
Abs	3	0.26	0.09	0.25
Abs	4	0.84	0.17	0.48
Abs	5	0.85	0.22	0.61
Abs	6	0.86	0.23	0.64

Table 207(Cont.): Judgment Type x Block (System Monitoring %Correct CONHL Blks 1-6)

JT	Block	Mean %Corr	Standard Error	Standard Deviation
Com	1	1.36	0.14	0.39
Com	2	1.04	0.16	0.44
Com	3	1.10	0.18	0.52
Com	4	1.26	0.16	0.44
Com	5	1.15	0.16	0.45
Com	6	1.36	0.14	0.39

Means, Standard Errors, and Standard Deviations for
Effects in Analysis on TLX

Table 208: Effect (n.s.): Judgment Type (TLX Overall Score)

Judgment Type	Mean OV	Standard Error	Standard Deviation
Absolute	56.98	6.48	18.33
Comparative	64.36	4.46	12.62

Table 209: Effect (n.s.): Condition (TLX Overall Score)

Condition	Mean OV	Standard Error	Standard Deviation
armhh	59.12	3.95	6.71
armlh	57.19	3.40	13.56
asmhh	59.25	4.24	16.97
asmhl	60.19	4.11	16.45
conhh	60.50	3.50	14.00
conlh	62.56	4.66	18.66
conhl	65.88	3.73	12.84

Table 210: Effect (n.s.): Judgment Type x Condition (TLX Overall Score)

JT	Cond	Mean OV	Standard Error	Standard Deviation
Abs	armhh	57.25	6.71	18.97
Abs	armlh	54.50	5.55	15.70
Abs	asmhh	54.12	6.95	19.66
Abs	asmhl	58.88	6.82	19.29
Abs	conhh	56.38	5.87	16.61
Abs	conlh	56.88	7.82	22.11
Abs	conhl	60.88	5.65	15.99

Table 210(Cont.): Effect (n.s.): Judgment Type x Condition (TLX Overall Score)

JT	Cond	Mean OV	Standard Error	Standard Deviation
Com	armhh	61.00	4.57	12.93
Com	armlh	59.88	4.03	11.43
Com	asmhh	64.38	4.62	13.06
Com	asmhl	61.50	2.83	14.28
Com	conhh	64.63	3.63	10.28
Com	conlh	68.25	4.79	13.54
Com	conhl	70.88	4.54	12.84

Table 211: Effect (n.s.): Judgment Type (TLX Frustration Score)

Judgment Type	Mean FR	Standard Error	Standard Deviation
Absolute	3.27	1.58	4.46
Comparative	2.02	1.13	3.21

Table 212: Effect (n.s.): Condition (TLX Frustrationl Score)

Condition	Mean FR	Standard Error	Standard Deviation
armhh	2.59	0.83	3.29
armlh	2.22	0.76	3.06
asmhh	2.07	0.72	2.86
asmhl	1.75	0.87	3.49
conhh	1.92	0.69	2.75
conlh	2.66	1.20	4.78
conhl	5.31	1.90	7.59

Table 213: Effect (n.s.): Judgment Type x Condition (TLX Frustrationl Score)

JT	Cond	Mean FR	Standard Error	Standard Deviation
Abs	armhh	3.35	1.39	3.93
Abs	armlh	2.39	0.88	2.49
Abs	asmhh	2.77	1.17	3.31
Abs	asmhl	2.50	1.66	4.70
Abs	conhh	3.00	1.22	3.45
Abs	conlh	2.29	1.30	3.68
Abs	conhl	6.62	3.43	9.67

Table 213(Cont.): Effect (n.s.): Judgment Type x Condition (TLX Frustrationl Score)

JT	Cond	Mean FR	Standard Error	Standard Deviation
Com	armhh	1.84	0.88	2.55
Com	armlh	2.04	1.31	3.71
Com	asmhh	1.38	0.83	2.34
Com	asmhl	1.00	0.58	1.65
Com	conhh	0.84	0.46	1.29
Com	conlh	3.04	2.10	5.93
Com	conhl	3.99	1.81	5.11

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level I SA in Res Mgmt Task: Res Mgmt vs Syst Monit

Table 214: Effect(n.s.): Judgment Type (Res Mgmt Level I Perception SA)

Judgment Type	Mean Level I SA	Standard Error	Standard Deviation
Absolute	0.79	0.18	0.52
Comparative	0.46	0.18	0.50

Table 215: Effect(n.s.): Condition (Res Mgmt Level I Perception SA)

Condition	Mean Level I SA	Standard Error	Standard Deviation
armhh	0.51	0.14	0.57
asmhh	0.73	0.12	0.48

Table 216: Effect(n.s.): Judgment Type x Condition (Res Mgmt Level I Perception SA)

JT	Cond	Mean Level I SA	Standard Error	Standard Deviation
Abs	armhh	0.62	0.22	0.63
Abs	asmhh	0.95	0.14	0.41
Com	armhh	0.39	0.18	0.52
Com	asmhh	0.52	0.17	0.47

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level I SA: Resource Mgmt vs Control

Table 217: Effect(n.s.): Judgment Type (Res Mgmt Level I Perception SA)

Judgement Type	Mean Level I SA	Standard Error	Standard Deviation
Absolute	0.57	0.17	0.49
Comparative	0.58	0.19	0.54

Table 218: Effect(n.s.): Condition (Res Mgmt Level I Perception SA)

Condition	Mean Level I SA	Standard Error	Standard Deviation
armhh	0.51	0.14	0.57
conhh	0.64	0.12	0.47

Table 219: Effect(n.s.): Judgment Type x Condition (Res Mgmt Level I Perception SA)

JT	Cond	Mean Level I SA	Standard Error	Standard Deviation
Abs	armhh	0.62	0.22	0.63
Abs	conhh	0.51	0.12	0.35
Com	armhh	0.39	0.18	0.52
Com	conhh	0.76	0.20	0.56

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level II SA in Res Mgmt Task: Res Mgmt vs Syst Monit

Table 220: Effect(n.s.): Judgment Type (Res Mgmt Level II Meaning SA)

Judgment Type	Mean Level II SA	Standard Error	Standard Deviation
Absolute	0.82	0.11	0.31
Comparative	0.66	0.15	0.42

Table 221: Effect(n.s.): Condition (Res Mgmt Level II Meaning SA)

Condition	Mean Level II SA	Standard Error	Standard Deviation
armhh	0.60	0.09	0.35
asmhh	0.88	0.10	0.41

Table 222: Effect(n.s.): Judgment Type x Condition (Res Mgmt Level II Meaning SA)

JT	Cond	Mean Level I SA	Standard Error	Standard Deviation
Abs	armhh	0.60	0.09	0.27
Abs	asmhh	1.04	0.13	0.36
Com	armhh	0.59	0.15	0.43
Com	asmhh	0.72	0.14	0.42

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level II SA in Res Mgmt Task: Res Mgmt vs Control

Table 223: Effect(n.s.): Judgment Type (Res Mgmt Level II Meaning SA)

Judgment Type	Mean Level II SA	Standard Error	Standard Deviation
Absolute	0.77	0.14	0.41
Comparative	0.66	0.11	0.30

Table 224: Effect(n.s.): Condition (Res Mgmt Level II Meaning SA)

Condition	Mean Level II SA	Standard Error	Standard Deviation
armhh	0.60	0.09	0.35
conhh	0.83	0.10	0.41

Table 225: Effect(n.s.): Judgment Type x Condition (Res Mgmt Level II Meaning SA)

JT	Cond	Mean Level II SA	Standard Error	Standard Deviation
Abs	armhh	0.60	0.09	0.27
Abs	conhh	0.94	0.19	0.55
Com	armhh	0.59	0.15	0.43
Com	conhh	0.72	0.06	0.16

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level I SA System Monitoring Task: Res Mgmt vs Syst Mon

Table 226: Effect(n.s.): Judgment Type (Syst Monit Level I Perception SA)

Judgment Type	Mean Level I SA	Standard Error	Standard Deviation
Absolute	1.16	0.15	0.58
Comparative	0.72	0.14	0.56

Table 227: Effect(sig.): Condition (Syst Monit Level I Perception SA)

Condition	Mean Level I SA	Standard Error	Standard Deviation
armhh	1.05	0.16	0.64
asmhh	0.83	0.14	0.58

Table 228: Effect(sig.): Judgment Type x Condition (Syst Monit Level I Perception SA)

JT	Cond	Mean Level I SA	Standard Error	Standard Deviation
Abs	armhh	1.37	0.20	0.56
Abs	asmhh	0.94	0.22	0.61
Com	armhh	0.72	0.20	0.56
Com	asmhh	0.72	0.20	0.56

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
Level I SA System Monitoring Task: Syst Mon vs Control

Table 229: Effect(n.s.): Judgment Type (Syst Monit Level I Perception SA)

Judgment Type	Mean Level I SA	Standard Error	Standard Deviation
Absolute	0.93	0.19	0.53
Comparative	0.64	0.18	0.52

Table 230: Effect(n.s.): Condition (Syst Monit Level I Perception SA)

Condition	Mean Level I SA	Standard Error	Standard Deviation
asmhh	0.83	0.14	0.57
conhh	0.74	0.12	0.49

Table 231: Effect(n.s.): Judgment Type x Condition (Syst Monit Level I Perception SA)

JT	Cond	Mean Level I SA	Standard Error	Standard Deviation
Abs	asmhh	0.94	0.22	0.61
Abs	conhh	0.91	0.16	0.46
Com	asmhh	0.72	0.20	0.56
Com	conhh	0.56	0.17	0.48

Means, Standard Errors, and Standard Deviations for Effects in Analysis on
Level II SA System Monitoring Task: Res Mgmt vs Syst Mon

Table 232: Effect(n.s.): Judgment Type (Syst Monit Level II Meaning SA)

Judgment Type	Mean Level II SA	Standard Error	Standard Deviation
Absolute	0.61	0.19	0.25
Comparative	0.58	0.07	0.20

Table 233: Effect(n.s.): Condition (Syst Monit Level II Meaning SA)

Condition	Mean Level II SA	Standard Error	Standard Deviation
armhh	0.65	0.06	0.25
asmhh	0.55	0.05	0.21

Table 234: Effect(sig.): Judgment Type x Condition (Syst Monit Level II Meaning SA)

JT	Cond	Mean Level II SA	Standard Error	Standard Deviation
Abs	armhh	0.72	0.10	0.29
Abs	asmhh	0.50	0.08	0.22
Com	armhh	0.57	0.07	0.21
Com	asmhh	0.60	0.06	0.18

Means, Standard Errors, and Standard Deviations for Effects in Analysis on Level II SA System Monitoring Task: Syst Mon vs Control

Table 235: Effect(n.s.): Judgment Type (Syst Monit Level II Meaning SA)

Judgment Type	Mean Level II SA	Standard Error	Standard Deviation
Absolute	0.53	0.09	0.24
Comparative	0.62	0.07	0.20

Table 236: Effect(n.s.): Condition (Syst Monit Level II Meaning SA)

Condition	Mean Level II SA	Standard Error	Standard Deviation
asmhh	0.55	0.04	0.20
conhh	0.61	0.06	0.24

Table 237: Effect(n.s.): Judgment Type x Condition (Syst Monit Level II Meaning SA)

JT	Cond	Mean Level II SA	Standard Error	Standard Deviation
Abs	asmhh	0.50	0.08	0.23
Abs	conhh	0.56	0.09	0.27
Com	asmhh	0.60	0.06	0.18
Com	conhh	0.65	0.08	0.22

APPENDIX H

GLOSSARY OF TERMS

Judgment Type

Absolute Judgment Type — one in which the desired state for the task must be maintained in memory; no visual or other referent available.

Comparative Judgment Type — one in which the desired state for the task is always available as a visual reference by which the current state may be compared.

Task Type

Internal Model — a collection of learned associative relationships/ patterns of associated events which reduce a pilot's/operator's search for information by directing attention away from irrelevant/ redundant cues/information.

Dynamic Internal Model — associated with a task in which the information relevant to decision-making (the relationships and properties within the task), once learned, do change across time.

Stable Internal Model	associated with a task in which the information relevant to decision-making (the relationships and properties within the task), once learned, do not change across time.
Resource (Fuel) Management	task associated with a dynamic internal model; the relationships between events and the meaning of events of the task change across time.
System Monitoring	task associated with a stable internal model; the meaning/relationships of the task do not change across time.

DISTRIBUTION LIST

	No. of Copies
Dr. Jim Ballas Naval Research Lab Code 5534 4555 Overlook Avenue, SW Washington, DC 20375-5000	1
LCDR John Deaton Defense Training and Performance Data Center 3280 Progress Dr. Orlando, FL 32826-3229	10
Dr. Peter Hancock 172 Pillsbury Dr. SE 164 Norris Hall Minneapolis, MN 55455	1
CDR S.D. Harris Naval Air Systems Command Code AIR-05TP3 Washington, DC 20361-3300	1
CDR J.M. Owens Naval Research Lab Code 5510 4555 Overlook Avenue, SW Washington, DC 20375-5000	1
Dr. Raja Parasuraman Dept. of Psychology Catholic University of America Washington, DC 20064	1
Dr. John Reising Cockpit Integration Directorate Code WRDC/KTC Wright-Patterson AFB, OH 45433	1
Dr. Joel Warm University of Cincinnati Dept. Of Psychology Cincinnati, OH 45221	1
Naval Air Warfare Center - Aircraft Division (2 for Code 8131)	2

DISTRIBUTION LIST (Continued)

	No. of Copies
Defense Technical Information Center ATTN: DTIC-FDAB Cameron Station BG5 Alexandria, VA 22304-6145	2
Center of Naval Analysis 4401 Fort Avenue P.O. Box 16268 Alexandria, VA 22302-0268	1
1299th Physiological Training Flight Malcolm Grow USAF Medical Center Andrews AFB, Washington, DC 20331-5300	1