ABSTRACT

WRIGHT, MELANIE CLAY. The Effects of Automation on Team Performance and Team Coordination. (Under the direction of David B. Kaber).

The advancement of technology has led to an increased use of automation in a number of work domains, including team environments. However, assessment of the effects of automation on teamwork has been primarily limited to the aviation domain (comparing early conventional aircraft models with more advanced aircraft cockpits) and studies have produced conflicting information regarding the impact of automation on team performance, communication, and coordination.

To more fully understand the implications of automation on system performance, researchers have begun to develop taxonomies and models of automation so that specific forms of automation can be defined and evaluated. A model proposed by Parasuraman et al. (2000) considers automation as it is applied to stages of information processing, including information acquisition, information analysis, decision selection, and action implementation. The objective of this research was to evaluate the effects of automation as applied to these different stages of information processing on the performance and coordination of teams in a complex decision making task.

A simulated Theatre Defense Task in which teams protect a home base from enemy attack was used as a test-bed for this evaluation. Two team members were required to work together to share information in order to successfully complete the task. One team member monitored incoming aircraft on a radarscope and used missiles to shoot down enemy aircraft. A second team member monitored information provided by reconnaissance aircraft to classify the incoming aircraft as enemy or friendly. Four automation conditions were designed that compared different degrees of information acquisition, information analysis, and decision selection automation. Two levels of difficulty, determined by the number of aircraft presented, were used in the experiment. Dependent measures for the experiment included team effectiveness, quantity of team communication, team coordination ratings by outside observers, and task and team workload ratings.

The results of the experiment revealed that different forms of automation have different effects on teamwork. Automation of information acquisition caused a decrease in the total amount of communication and an increase in the ratio of information transferred compared to information requested between team members. Automation of information analysis resulted in higher team coordination ratings. Automation of decision selection led to better team effectiveness under low levels of task difficulty but at the cost of higher workload. The fact that differing forms of automation had different influences on team performance in this research aids in explaining conflicting historical findings regarding the effects of automation on teamwork. The results of this research may have utility for the design of complex systems used in team environments.

THE EFFECTS OF AUTOMATION ON TEAM PERFORMANCE AND TEAM COORDINATION

By

MELANIE CLAY WRIGHT

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

DEPARTMENT OF INDUSTRIAL ENGINEERING, NCSU

Raleigh

2002

APPROVED BY:

_______________________________ _______________________________

_______________________________ _______________________________

Chair of Advisory Committee

DEDICATION

To my nieces and nephew, Hannah, Christine, Julia, Harrison, and Miranda. I hope your education, in whatever form it takes, helps you to find a lifetime of happiness.

BIOGRAPHY

Melanie Clay Wright was born Melanie Carol Clay in Bethesda, Maryland in April, 1966. She was raised in Rockville, Maryland, where she completed her elementary and secondary education, graduating from Magruder High School in 1984.

Ms. Wright studied Aerospace Engineering at Virginia Polytechnic Institute and State University in Blacksburg, Virginia. Following graduation from Virginia Tech in 1988, she worked as an aerospace engineer at the Naval Aviation Depot in Cherry Point, North Carolina for two years.

In the fall of 1990, the author began studies in Psychology (Ergonomics) at North Carolina State University in Raleigh North Carolina. She received her M.S. in Psychology in 1994. Immediately following her M.S. degree, Ms. Wright began the Ph.D. program in Industrial Engineering (Ergonomics).

In 1994, the author married Innes James Wright. They have two golden retrievers, Bob and Lucy.

From 1992 to 2000, Ms. Wright was employed with Monterey Technologies, Inc. in Cary, NC. From 2000 to 2001, she was employed with Ericsson, Inc. Ms. Wright is currently employed by North Carolina State University and will begin a new job with SA Technologies, based in Marietta, GA upon completion of her degree.

ACKNOWLEDGEMENTS

I wish to thank a number of people for their assistance and support in the course of this work. First I would like to thank my advisor and committee chair, Dr. David Kaber. Dr. Kaber not only provided me with the opportunity and support to finally complete this effort, but also provided guidance with all aspects of my work so that the quality of the final product was raised to a standard of which we can both be proud. In this past year, Dr. Kaber has reminded me of the meaning of lifelong learning and the importance of scholarly work.

I would also like to extend a special thank you to Srikanth Jayanthi for his advice and incredible effort watching nearly 100 hours of what must have been excruciatingly boring video tape as a team rater.

My sincere thanks the members of my committee Dr. Sharolyn Converse Lane, Dr. Michael Kay, and Dr. Gary Mirka for their advice and encouragement. To Dr. Lane and Dr. Kay, thank you for sticking with me through the long haul and reviewing early versions of proposals, some that made it this far and some that did not. I appreciate your enthusiasm and flexibility in supporting all of my ideas. Thank also to my graduate school advisor, Dr. Susan Osborne for her contributions and interest.

Thanks to Cheryl Bolstad, Mark Bolstad, and Mica Endsley for their assistance in providing me with software and guidance regarding the Theater Defense Task for my experiment.

My thanks go to the following friends and family for supporting me by participating as subjects and/or by acting as great recruiters:

Janey Barnes, Teri Brooks, Megan Casey, Naomi Glasscock, Elizabeth Guertin, Michelle Khoury, Karen Lloyd, Bonnie McCullough, Grant McCullough, Karen Noack, Bobby Noell, Amy Pfeiffer, Melinda Pfeiffer, Laura Taillon, Tim Taillon, Hester May Stols, Heather Warren, Innes Wright, Greg Young.

I would also like to my mother-in-law, Betty Wright; sisters-in-law, Zane and Sandra Wright; and “The Girls”. These intelligent, confident, and caring women have been supportive role models for me over the past the ten years.

Finally I would like to thank my three biggest "fans". To my mother, Mary Clay, thank you for your persistent and gentle encouragement to keep on going until everything was finished and for the upbeat phone calls that always lifted my spirits. I know you would have been a subject 40 times and would have typed the entire manuscript for me if you could have. To my father, Leon Clay, thank you for setting an example for me to always try to set my sights just a little higher, both in my education and in life. I know that you are always quietly cheering for me on the sidelines. To my husband, Innes Wright, thank you for supporting me in countless ways, but most of all thank you for helping me to keep my perspective and for making me laugh every day.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The advancement of computer technology has lead to an increased use of automation in a wide variety of work domains. These systems are capable of performing tasks that have previously been the realm of human operators. Automation is prevalent in aircraft cockpits, military applications, medical environments, power plant applications, and personal computers. A large number of work environments that use automation are so complicated that they require multiple operators (Bowers et al., 1996). In many of these environments, interaction between human operators and automation is critical for system safety.

Researchers have noted human-machine interaction problems related to increasing levels of automation in aviation and other domains (Bainbridge, 1987; Coury and Semmel, 1996; Sarter, 1994; Woods, 1996). They have also proposed a number of solutions to the problems imposed by increasing automation (Bainbridge, 1987; Coury and Semmel, 1996; Parasuraman et al., 1996; Scerbo, 1996). There is reason to believe that problems also will exist in team environments. A number of researchers have suggested that automation may qualitatively change the communication between human team members (Cannon-Bowers, Salas, and Converse, 1993; Johannesen et al., 1994; Wiener, 1993).

Interactions between human operators in a team situation are often sub-optimal. Aviation accident and incident data indicate that 70% - 80% of accidents and incidents are in part attributable to "human error" (Billings and Reynard, 1984; Helmreich and Foushee, 1993). Similar numbers (75% to 80%) of anesthesiology mishaps in the medical field have been attributed, in part, to human error (Helmreich and Schaefer, 1994). Further analysis of the aviation data indicates that most of the "human errors" are due to communication problems (Billings and Reynard, 1984; Helmreich and Foushee, 1993). Since communication is such a well-known problem in complex systems, it is important to understand how automation affects communication within the team. An entire field of research is devoted to the study of crew resource management (CRM) and team communication in work environments (Wiener, Kanki, and Helmreich, 1993). However, very little research has evaluated team performance and communication in work environments that include automated systems (e.g., Bowers et al., 1993; Costley, Johnson, and Lawson, 1989). This dissertation describes a research study to evaluate the effects of automation on team coordination and team performance.

The following literature review is organized into three main sections. The first section covers automation, detailing different forms of automation and methods of characterizing and categorizing automation. It also covers advantages of automation, problems related to automation, and potential solutions to the problems. The second section covers research related to teams, including measurement of teams, and characteristics of high performing teams. The final section reviews research conducted on the effect of automation on the performance and communication patterns of teams.

2. AUTOMATION

2.1. Terms and Definitions

Automation comes in many different forms. A number of terms including automation, intelligent agents, expert systems, and decision-aiding systems have been used to describe various forms of automation. According to Parsons (1985), automation can be thought of as the process of allocating activities to a machine or system to perform. However, practitioners usually take this definition a step further and consider automation in relation to current technology. That is, automation can be thought of as the process of allocating activities to a machine or system, which have been performed by humans in the recent past. Billings (1997) defines automation as a system or method "in which many of the processes of production are automatically controlled by autonomous machines or electronic devices." Billings (1997) views automation as a "tool" that can allow a human operator to accomplish tasks that would otherwise be difficult or impossible, or to carry out actions independently that would otherwise require increased human attention or effort. According to Sarter and Woods (1997), "Automation refers to a wide variety of systems that differ with respect to their capabilities and design features" (Sarter and Woods, 1997). Clearly there is a wide range of system characteristics and capabilities that may be classified as automation.

Examples of automated systems include automobile cruise control, autopilots in aircraft, and automated systems monitoring. Early aircraft autopilots could hold an aircraft straight and keep the wings level. More advanced technology cockpits such as the Airbus A-320, the Boeing 757/767/737-300, and the MD-88 have automation capabilities that include integrated flight guidance systems, automated control surfaces, and aircraft system monitors (Wiener, 1993).

The term "intelligent agent" is used, rather than automation, when the system exhibits two characteristics including (1) "intelligent" implies that the system has the capability to reason about a task and learn from task performance (Chen, et al., 1996); and (2) "agent" generally implies that the assistance is presented to the user in some anthropomorphic or otherwise expressive form (Milewski and Lewis, 1997). According to Chen et al., an intelligent agent is: (1) integrated (into a consistent user interface), (2) expressive, (3) goal-oriented, (4) cooperative, and (5) customized (to different users).

Most intelligent agent examples are in the form of personal computer software applications. Maes (1994) describes four personal data assistant agents that learn from the user by observing and imitating the user, receiving feedback from the user, receiving explicit instructions from the user, and by asking other agents for advice. Lester and his colleagues (Lester et al., 1997) have developed intelligent agents such as these for use in intelligent learning environments.

The terms "expert system" and "decision aiding system" also have a similar definition as "automation" and "intelligent agents". According to Sheridan and Thompson (1994), an expert system is a computer-based system, with its associated knowledge base and algorithms, that can draw conclusions and give advice on a particular subject. Examples of expert systems include MYCIN, a medical diagnostic aid (Card, 1989), flight planning aids (Layton, Smith, and McCoy, 1994), and the Pilot's Associate, an aircraft systems and event monitoring and planning system (Hammer and Small, 1995).

2.2. Levels of Automation

It is difficult to draw distinctions between different types of automated systems. Often, different terms are used interchangeably to describe a single system. Or, a system may have characteristics that categorize it into more than one of the three types of systems described (intelligent agents, expert systems, and decision-aiding systems). For example, even though the term "intelligent agent" is generally used when a system learns and adapts over time, systems referred to as "expert systems" may also learn and adapt over time.

Some researchers, rather than distinguishing through the use of terms like automation and intelligent agents, refer to different levels of automation (LOAs) when discussing these types of systems. Scerbo (1996) reviewed different categorizations of levels of automation by various researchers. Sheridan and Verplanck (1978) developed a ten level taxonomy of levels of automation that distinguished between such characteristics as the number of decision alternatives the system provides the user and the amount of information provided from the system to the user. Wickens, et al. (1998) modified this taxonomy slightly to present the ten levels representing high to low automation shown in Table 1. However, this taxonomy really only applies to automation of the decision-making part of an operator's task. Wickens et al. (1998) suggest that the levels of automation can range across three different scales: (1) Information Acquisition and Integration from high to low; (2) Decision and Action Selection from high to low; and (3) Action Implementation as either automatic or manual.

Table 1. Levels of automation (Wickens et al., 1998).

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

Considering a wide range of domains (e.g., aviation, air traffic control, advanced manufacturing, and teleoperations) that require an array of cognitive and psychomotor tasks, Endsley and Kaber (1999) developed a similar taxonomy based on four generic information processing functions that could be automated. These functions are:

(1) monitoring – scanning displays to perceive system status,

(2) generating – formulating options or strategies for achieving goals,

(3) selecting – deciding on a particular option or strategy, and

(4) implementing – carrying out the chosen option.

Endsley and Kaber (1999) established ten unique levels of automation based on the assignment of these functions to either the human operator or the computer. The taxonomy is shown in Table 2.

Table 2. Taxonomy of levels of automation applicable to dynamic-cognitive and psychomotor control task performance (Endsley and Kaber, 1999).

Level of Automation	Roles			
	Monitoring	**Generating**	**Selecting**	**Implementing**
1. Manual control	Human	Human	Human	Human
2. Action support	Human/Computer	Human	Human	Human/Computer
3. Batch processing	Human/Computer	Human	Human	Computer
4. Shared control	Human/Computer	Human/Computer	Human	Human/Computer
5. Decision support	Human/Computer	Human/Computer	Human	Computer
6. Blended decision-making	Human/Computer	Human/Computer	Human/Computer	Computer
7. Rigid system	Human/Computer	Computer	Human	Computer
8. Automated decision-making	Human/Computer	Human/Computer	Computer	Computer
9. Supervisory control	Human/Computer	Computer	Computer	Computer
10. Full automation	Computer	Computer	Computer	Computer

More recently, Parasuraman et al. (2000), expanded Wickens' scales of automation into four categories of human information processing that coincide with the four functions presented by Endsley and Kaber (1999) – (1) Information Acquisition, (2) Information Analysis, (3) Decision Selection, and (4) Action Implementation. Parasuraman et al. (2000) state that for each of these functions, the level of automation can be classified from low to high. For example, the scale presented in Table 1 may represent the classification of the Decision Selection phase of processing into levels of automation from low to high.

Classifications of levels of automation such as these serve to qualitatively describe the characteristics of automated systems so that they may be evaluated and compared in scientific research. Now, for example, rather than distinguishing through terms such as automated systems, intelligent agents, and decision-aids, we may describe systems based on how much automation is provided for each step within the processing of a task. A traditional "decision aiding system" may be represented by a function that provides automation of information acquisition and information analysis, but leaves decision selection and action implementation up to the human operator.

2.3. Advantages, Problems, and Solutions of Automation

2.3.1. Advantages

The advantages of automation are fairly obvious. Automation is intended to lighten the workload for humans, and, support system operations with fewer errors. While humans are generally poor at monitoring or vigilance tasks, computers are good at performing these types of tasks (Endsley, 1996; Mosier, and Skitka, 1996). Because computers can process more information and process information more quickly than humans, expert systems or automated decision aids can provide advice that takes into account more information than a human decision maker is likely to consider. Humans are notoriously poor at considering all of the factors involved in making decisions. They generally consider only a few inputs and the combination of these inputs tends to be additive (Brehmer, 1987).

With the assistance of automation, systems may be operated with fewer operators. According to Wiener (1993), "Automation was to play a major part in the emerging controversy over the two- versus three-pilot crew. With the growing sophistication of cockpit automation, there seemed less and less justification for the position of a flight engineer." In fact, in 1981, a President's Task Force on Aircrew Complement decreed that modern airliners, including wide-bodies could be flown with a crew of two, unaided by a flight engineer. Computers could perform many of the functions previously performed by the flight engineer, such as systems monitoring.

2.3.2. Automation Problems

Although automated systems clearly present advantages, research and experience with these systems indicate a number of potential problems. Funk, Lyall, and Niemczyk (1997) created a taxonomy of 85 different problems and concerns with the use of automation from the Aviation Safety Reporting System (ASRS) accident and incident database. Fifty of these problems and concerns were verified through a survey of flight deck automation experts. Common automation problems include:

- poor feedback from the system,

- mode errors or mode confusion on the part of the user,
- complacency or over reliance on the automation by the user,
- loss of skill required of the user to perform the task manually, and
- poor performance by the user in abnormal conditions.

Poor feedback or lack of feedback from the system to the user makes it difficult for the user to understand how the automated system is functioning. One reason automated systems fail to adequately inform the users of their actions is that the automated systems place an additional layer of complexity (data processing, data fusion, and intelligent control) between the actual system processes and sensory data the user is controlling (Coury and Semmel, 1996). In addition, this additional layer of complexity may include artificial intelligence and engineering analyses that the intended users don't understand. Some type of translator must exist to provide understandable feedback to the user.

Woods (1996) discusses the problem of poor feedback in automated systems. Woods states that the amount of data available to the human is increasing. However, the effectiveness of that data depends on the cognitive work required for the human to turn it into a coherent interpretation in context. Even though systems may present all of the necessary data, they exhibit "low observability" in that the data is presented in such a way that too much cognitive work is required to interpret it. Woods indicates that this is especially problematic when systems also exhibit a high degree of autonomy and authority; that is, "strong but silent" automated systems may have latent dangers. For example, in the Airbus A-320, the automation sets different priorities with respect to maintaining speed and path (based on fuel efficiency) that result in a deviation from the target altitude that is different from how human pilots would fly. This behavior is not made apparent to the pilot and, thus, results in surprises for the pilot (Sarter, 1994).

Complex automated systems (such as flight deck automation) often operate in a number of different modes in which the behavior of the system is slightly different for each mode. Users often find themselves unaware of the current mode of the system and, as a consequence, surprised by the system behavior (Sarter and Woods, 1997). Andre and Degani (1997) discuss mode awareness and describe problems of mode awareness associated with relatively common systems such as remote controls and automobile cruise controls. In these cases, users have problems because they forget what mode they have set (such as "VCR" vs. "TV", or cruise control "on") and don't understand when the system reacts differently than they expect.

In the case of more complex systems such as flight deck automation, the transition between modes may occur automatically, making it even more difficult for the crew to remain aware of the current mode of the system (Sarter and Woods, 1995). Aviation accident and incident data indicate that pilot behavior appropriate for one mode, when a different mode is active, is a common problem leading to accidents or incidents (Sarter and Woods, 1995; Wickens et al., 1998). The problem of mode errors is related to the problem of poor feedback. Systems that do not clearly convey the current mode of operation fail to provide adequate feedback to the user.

Parasuraman et al. (1996) cite data from the ASRS indicating that pilot over-reliance on automation is thought to be a contributing factor in aircraft accidents and incidents. These incidents generally involve a probable failure in monitoring on the part of the crew. Complacency, or over-trust in an automated system may cause users to miss automation failures, especially if failures are rare (Wickens et al., 1998). Trust in automation has been studied by a number of researchers (Lee and Moray, 1992; Muir, 1987; Riley, 1994, 1996). Just as too much trust can lead to complacency in the monitoring of failures, too little trust or mis-trust may lead to a failure to use automation when it is appropriate to do so. Therefore, operators may miss out on beneficial reductions in workload that may be provided by automation. Riley (1994, 1996) has studied operator reliance on automation and the factors that influence use of automation. He found that performance uncertainty (or difficulty, independent of workload), perceived risk, and automation accuracy (reliability) affect the operator's decision to rely on an automated system. Riley also found that the decision to rely on automation and the factors that affect this decision are largely subject to individual differences among operators.

The loss of skill in operating a system manually is another common problem associated with advanced automation. According to Bainbridge (1987), "…physical skills deteriorate when they are not used, particularly the refinements of gain and timing." This type of skill loss is problematic in tracking tasks such as driving, flying, or target tracking. In addition, cognitive skills such as the efficient retrieval of knowledge from long-term memory are also dependent on the frequency of use of the information (Bainbridge, 1987). The implication of this skill loss is that when an operator is forced to take over an automated system manually, he or she is likely to do so with minimal information and skill. A recent study using an MD-11 simulator showed that pilots in a supervisory control condition (using an automated Flight Management System) were less complete in their flight planning than pilots in manual control conditions and, thus, would be less prepared to take over if necessary (Kaber et al., 2002).

The problem of skill loss leads directly to the final problem of automation to be discussed. Sarter (1991) found that operators appeared to perform effectively with automated systems during normal conditions, but during abnormal conditions, operators did not always react appropriately to the failure of automated systems. This has been referred to as the "out-of-the-loop performance problem" or the "out of the loop unfamiliarity problem" (Endsley and Kiris, 1995; Kaber and Endsley, 1997a; Wickens and Hollands, 2002). There are several possible reasons for this problem. Loss of skill, poor feedback, and complacency may all contribute to the out-of-the-loop problem. Another likely reason for the out-of-the-loop problem is that, because the system is under automated control, the users are not actively involved in controlling the system. Therefore, they are less likely to detect malfunctions (Parasuraman et al., 1996) and are less likely to be prepared to take over when malfunctions do occur (Bainbridge, 1997). Endsley and Kiris (1995) found that subjects under conditions using automation for a decision-making task performed worse in answering situation awareness questions related to understanding of the system (more in-depth knowledge than simply knowledge of actions taken) than did subjects in manual conditions. They attributed this effect to "the difference between active and passive processing of information".

Bainbridge (1997) proposes one other reason for many of the problems associated with automated systems. She suggests that designers automate the easy parts of a human operator's task, leaving the operator to do the tasks that the designer is unable to automate. The operator is then left with an arbitrary collection of tasks, with little thought given to providing support for the operator.

2.3.3. Solutions to Automation Problems

Researchers have proposed a number of solutions to problems with automation. Unfortunately, without analyzing specific systems in detail, these proposals come in the form of general recommendations. Several studies have recommended a human-centered design process for the design of automated systems (Bainbridge, 1987; Billings, 1997; Endsley and Kaber, 1999; Kaber, 1997; Wiener, 1993). The design of systems should be based on the needs of the operators rather than driven by the available technology (Wiener, 1993). Wiener states that in addition to properly designing hardware and software, a human-centered design will include support documentation, procedures, checklists, operational doctrine, and training programs. Bainbridge (1987) suggests that methods of man-computer collaboration need to be more fully developed. She recommends systems that assist in decision-making by instructing or advising the operator, by mitigating operator errors, by providing sophisticated displays, and by assisting the operator when task loads are high. Kaber and Endsley (in review) note a number of important aspects of human-centered automation including adequate feedback, predictable functioning, support of operator achievement of situation awareness, and assignment of tasks to the human and computer such that a team effort is achieved.

A number of researchers have proposed a program of adaptive automation to counteract problems of automation such as loss of skill and complacency. In adaptive automation, the level of automation, or the number of systems monitored, is modified in real time (Scerbo, 1996). Both the human and machine may share control over changes in the state of automation. Adaptive automation is intended to promote an optimal coupling of automation to operator workload and situation awareness (Kaber and Riley, 1999; Kaber et al., 2001). When workload is low, the human may control more of the system, maintaining skill and awareness; however, when workload is high, the machine takes over some tasks for the operator. Parasuraman et al. (1996) have studied methods for controlling task allocation between the man and machine based on a model of expected task difficulty or operator performance in a task. Prinzel et al. (1995) demonstrated that physiological measures (EEG data) could serve as a basis for directing the operator's level of engagement in a task. Kaber and Riley (1999) demonstrated that embedded secondary task measures of workload could also be used as a basis for triggering adaptive automation to maintain operator workload at desired levels.

Based on the classification of automation into different levels across different functions (e.g., see Table 2), researchers have suggested identifying appropriate levels of automation as a solution to some of the previously discussed problems of automation (Endsley and Kaber, 1999; Endsley and Kiris 1995; Kaber and Endsley, 1997a, 1997b; Parasuraman et al., 2000). Similar to the goals of adaptive automation, Endsley and Kaber (1999) have found that providing the right degree of automation for the right function of the task can optimize the use of automation for operator workload and situation awareness. They found that levels of automation that combine human generation of options with computer implementation of actions produced better overall performance during normal operations of a laboratory simulation of an air traffic control task (Endsley and Kaber, 1999). Kaber et al. (1999), using a high fidelity simulation of a telerobot, found that high levels of automation enhanced performance and reduced workload during normal operation conditions. Intermediate levels of automation promoted higher operator situation awareness and enhanced manual performance during system failure modes as compared to higher levels of automation. Ruff et al. (2000) compared three levels of automation (manual, management by consent, and management by exception) across two levels of automation reliability in a simulation of the control of remotely operated vehicles. They found performance and situation awareness advantages to the management by consent condition over the manual and management by exception conditions.

Kaber and Endsley (1997b) looked at the effects of both level of automation and adaptive automation on performance, workload, and situation awareness in a laboratory simulation similar to an air traffic control task. They found that both level of automation and adaptive automation affected system functioning, but in slightly different ways. Level of automation had a significant effect on performance of the main control task while adaptive automation affected performance in a secondary monitoring task that indicated an impact of operator workload. They found the best combination of level of automation and adaptive automation to be human strategizing of options with computer implementation of actions under high adaptive automation cycle times (longer periods of automated performance between manual control cycles). Based on this research, Kaber and Endsley (1997b) were able to conclude:

> "Although LOA and AA have been found to significantly interact in effecting performance, changes in the allocation of function to human and computer servers (i.e., LOA) appears to be far more important to performance improvements than the amount of time that is spent on a task under automated versus manual control."

Another proposal for dealing with the problems of automation is to design systems with better feedback and communication with the operator (Bainbridge, 1987; Coury and Semmel, 1996). Coury and Semmel state that one of the primary goals of the user interface is to direct the user's attention to information to the decision-making process for a specific situation. Bainbridge (1987) suggests that the design of displays may assist in problems with automation by presenting operators with information such as target values for monitoring systems, versus simply providing alarms when systems operate outside a target range. In addition, Bainbridge states that because automatic control can camouflage system problems, that feedback regarding automatic system failure must be obvious. Sarter (1996) states that while many researchers call for salient indications of automation failures or automation state changes; little research has been conducted to determine what makes an indication salient. Sarter (2000) has followed through on this observation by exploring different forms of sensory displays and found that the distribution of information across various modalities leads to improvements in the use of automation, related to mode awareness problems. In particular, Sarter has found that using peripheral visual cues and tactile cues in addition to the more common foveal visual displays and auditory feedback is helpful.

A final suggestion by researchers for dealing with automation problems is to design systems that improve communication between the automated system and the operator by viewing automation as a team member (Bowers et al., 1996; Scerbo, 1996; Woods, 1996). Woods (1996) states that the design of automated systems is actually the design of a team that requires provisions for coordination between machine agents and the operators. Bowers et al. (1996) suggest that automated systems may serve as a convenient "team member" taking on tasks with no negotiation during high workload to help optimize team resources. However, depending on the implementation, this may conflict with problems raised by Woods (1996) of "low observability". Johannesen, Cook, and Woods (1994) suggest that there should be a "common frame of reference" or common accessibility of the problem state and problem solving approach between the human and the machine. Scerbo (1996, 2000) states that understanding human team dynamics should be the force guiding the development of adaptive automation. He references the conclusions of Hammer and Small (1995) that research on how humans share tasks and information would have been more useful in designing their Pilot's Associate. Scerbo (2000) states that it is not necessary to have the technology currently available to understand how humans interact with adaptive technology, rather, "one can look at how humans share tasks for guidance along these lines."

2.4. Summary

This section has defined automation as the process of allocating tasks to machines that have, in the recent past, been performed by humans. Subtleties in the definitions of automation, intelligent agents, and expert systems have been described that help to explain the use of different terms. Unfortunately, these definitions and the slight differences between them do not provide objective criteria to distinguish systems so that we may associate effects on human-system performance to specific types of systems. As an alternative, an approach defining automation based on levels of automation assigned to different functions of a task has been presented. This theoretical organization of complex systems will help us to link effects on human-system performance to specific characteristics of automation – from low to high levels of automation across the functions of information acquisition, information analysis, decision-making, and action implementation.

This section has also presented problems that have been associated with automation. Problems that automation poses for operators include poor feedback, mode errors, complacency, loss of skill, and out-of-the-loop performance problems. Solutions to these problems include improved feedback (particularly display design), adaptive automation, and allocation of levels of automation based on the function of the task. One additional solution – observing human-human interaction to identify appropriate sharing of tasks and information between the human and computer – has been proposed but not evaluated.

Most of the research discussed so far has been in the context of single-operator systems. However, many of the work environments that use automation require two or more operators interacting with a system to complete the task at hand. Examples include advanced aviation, medical treatment, and power plant operation. The next section presents a definition of teams, describes theories of team performance, presents research with respect to measuring teams and discusses characteristics of high- versus poor- performing teams. The main reason for reviewing research related to teams is to understand the current theories and accepted measures of team research in order to design and conduct an experiment to assess the effects of automation on teams.

3. TEAMS

3.1. Definition

A definition of a team was proposed by Salas, et al. (1992):

"…a *team* is defined as a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership."

Salas et al. point out that this definition implies that task completion requires: (1) a dynamic exchange of information and resources between team members; (2) coordination of task activities; (3) adjustment to task demands; and (4) organizational structuring of members. The interdependence characteristic of this definition indicates that the output of one member's task is a critical input factor for another member's task (Bowers et al., 1996). Brannick and Prince (1997) state that coordination is a central feature of teamwork. According to Brannick and Prince, coordination implies either simultaneity (team member tasks must be coordinated to be completed at the same time) or sequencing (team member tasks must be coordinated to be completed in a specific order).

Salas et al. (1992) also argue that teams can be conceived to fall on a continuum with highly structured, interdependent teams at one end of the continuum and teams whose members interact minimally and perform individual tasks in a group context at the other end.

3.2. Theories of Team Performance

Theories of team performance assist researchers in understanding the factors behind team performance and help researchers follow an integrated approach to designing research studies of the performance of teams. Salas et al. (1992), reviewed several models of team performance and evaluated the research that supported these models to formulate a single "Integrated Model of Team Performance". More recently, this model has been supported and presented as the “Team Effectiveness Model (TEM)” (Weaver et al., 1995; Bowers et al., 1996). Figure 1 presents this model.

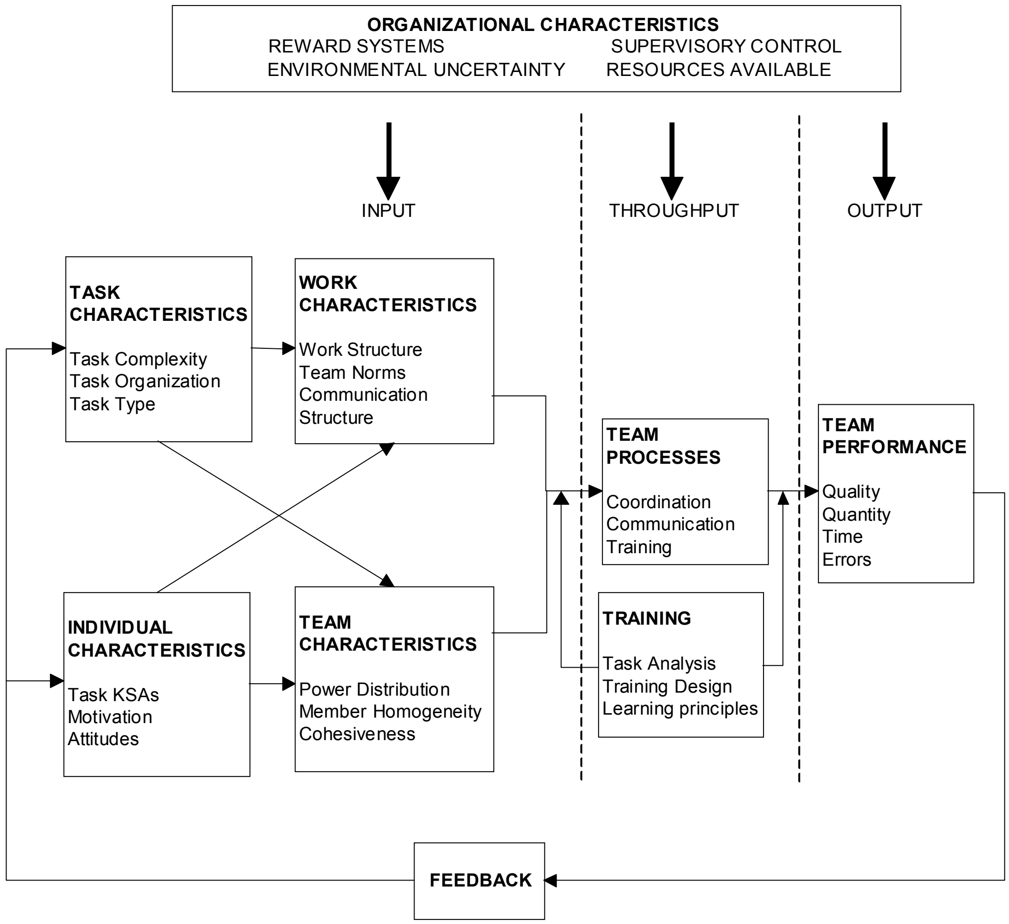

Figure 1. The Team Effectiveness Model (TEM). Adapted from Salas, Dickenson, Converse, and Tannenbaum (1992).

The team effectiveness model specifies the stages related to team performance as input, throughput, and output (Weaver et al., 1995). Inputs include task, work, individual, and team characteristics. Examples of these characteristics include complexity of the task, work structure, individual skills, and cohesiveness of the team. Throughputs are the processes that the team uses to produce performance outputs. These include team training, communication, and coordination. Finally, the outputs of the team effectiveness model are the quantity and quality of work or products by the team. Organizational characteristics such as available resources provide a context that influences all three stages of the model.

Feedback on team performance feeds into individual and task characteristics to moderate the processes and output of the team over time.

Helmreich and Foushee (1993) used a similar framework for their model of flight crew performance in an evaluation of crew resource management. Their model includes input factors such as individual aptitudes and the operating environment, group process factors such as communication and decision strategies, and output factors that include mission and crew performance. Studies of team performance may evaluate the influence of one or more input factors on both the throughput and the subsequent output of the team.

3.3. Measuring Teams

Following the Team Effectiveness Model, there are a number of ways in which to measure teams. Teams may be measured on the basis of their output through measures of task performance in terms of both quantity and quality. In addition, measures of the throughput or team processes may be gathered. Based on these types of measures, conclusions can be made as to the relationship between input characteristics, team processes, and team performance.

3.3.1. Team Effectiveness Measures

Kraiger and Wenzel (1997) refer to measures of the outcome of teams as team effectiveness measures. The term *team effectiveness* is used, rather than the term *team performance*, because the term *team performance* has been used interchangeably in literature to describe both the outcome measures of a team (e.g., as shown in the TEM in Figure 1) and measures of how a team carries out tasks, or process measures (Kraiger and Wenzel, 1997). The measurement of team effectiveness is usually straightforward. As with measures of individual performance, depending on the task to be completed, there are usually objective measures that can be recorded. These measures may include time on task, quantity of work completed, number of correct responses, number of errors, etc. The measures may be scored separately for individuals on the team and as team totals.

3.3.2. Team Process Measures

Measures of team processes have been organized and defined along a number of different dimensions. According to Brannick and Prince (1997), process measures may focus on interpersonal stylistic elements or on task-oriented elements of the process. For example, task-oriented measures may focus on details such as how long a problems goes unnoticed or which team member provides corrective action; while a measure that focuses more on interpersonal style would focus more on how the corrective action was given and accepted by the team members. A number of researchers have developed taxonomies or listed processes that are important to effective teamwork. Table 3 presents a small sample of these taxonomies. However, no consensus exists for a specific set of team process measures that describe and differentiate the performance of teams along constructs such as those listed in Table 3 (Brannick and Prince, 1997; Kraiger and Wenzel, 1997). One obvious reason for the lack of consensus is that the processes essential to teamwork or team performance are likely to be task-dependent. Researchers evaluating teams in different task environments are likely to identify different team process factors as important.

There does appear to be a consensus, however, that the measurement of team process, in addition to the measurement of outcomes, is important in the study of teams. Brannick and Prince (1997) describe a sailing team as an example. The outcome of the team in terms of time to complete the race can be due to factors other than simply teamwork, for example, the characteristics of the boat. By measuring team processes, we may have better insight into the functioning of the team than if we restricted ourselves to outcome measures alone. In addition, team process measures allow for the study of the team throughout the task, assessing teamwork at different points in time, rather than strictly limiting assessment to the final result.

Table 3. Sample of proposed team process constructs.

Reference	Proposed Team Process Constructs	
Morgan et al., 1986	• Communication • Cooperation • Team Spirit • Giving suggestions	• Acceptance of suggestions • Coordination • Adaptability
Bettenhausen, 1991	• Cohesion • Commitment	• Conflict • Goal setting
Fleishman and Zaccaro, 1992	• Orientation • Resource distribution • Timing (activity pacing) • Response coordination	• Motivational • Systems monitoring • Procedure maintenance
Prince, et al., 1992 Aircrew Coordination and Observation Scale (ACOE)	• Communication • Situation awareness • Leadership • Assertiveness	• Decision-making • Mission analysis • Adaptability
Helmreich and Foushee, 1993	• Communication and decision behavior • Team building • Workload management and situation awareness • Operational integrity	
McIntyre and Salas, 1995	• Decision-making • Communication • Leadership • Coordination • Adaptability/flexibility	• Assertiveness • Situational awareness • Morale • Feedback • Backup behaviors
Dickinson and McIntyre, 1997	• Team orientation • Team leadership • Communication • Monitoring	• Feedback • Backup behavior • Coordination

3.3.2.1. Team Communication Measures

One team process that is identified in six of the seven taxonomies summarized in Table 3 is communication. Communication among team members has been measured in a number of studies of team performance. One method of measuring communication is to first categorize types of communications, then record the number of communications that fall into each category. Costley, Johnson, and Lawson (1989) classified communication into the following categories:

- Commanding
- Reacting
- Information processing
- Giving explanation
- Checking

- Summarizing
- Asides
- Questioning
- Seeking information
- Testing understanding

Kanki and Palmer (1993) used a much shorter list of communication categories including (1) commands, (2) questions, and (3) acknowledgments. Urban et al. (1993) coded types of communication into the following categories:

- Questions
- Answers
- Answers which formed requests
- Responses to requests
- Statements
- Acknowledgments of prior utterances

Another measure that has been explored with respect to communication is a comparison between explicit and implicit forms of communication (Serfaty, Entin, and Volpe, 1993). In measuring implicit communication, methods of communicating other than speech such as gestures or touch (e.g., pointing to a display or tapping on the arm) are recorded. Serfaty, Entin, and Johnston (1998) suggest that implicit coordination can also be measured by considering the ratio of information transfers to information requests. They define this measure as the *anticipation ratio*. Macmillan, Entin, and Serfaty (in press) assert that teams with a higher anticipation ratio are more effective than teams with a lower anticipation ratio.

3.3.2.2. Team Coordination Measures

The measurement of other team processes that are indicated in Table 3 is often referred to as the measure of team coordination. Constructs such as decision-making and assertiveness are difficult to measure using objective measures. Generally the measurement of team coordination, beyond that of communication, has relied on either team or external observer ratings of team coordination. Tesluk et al. (1997) summarize a number of research studies that have used surveys, observations, interviews, or archival data reported by incumbents, subordinates, peers, supervisors, or external experts.

Bowers et al. (1992) recommend, and Jentsch et al. (1995) used, the Aircrew Coordination and Observation Scale (ACOE) to assess the frequency and quality of coordination behaviors along seven behavioral objectives including communication, situation awareness, leadership, assertiveness, decision-making, mission analysis, and adaptability. In this scale, behaviors such as acknowledging communication, asking for clarification, using standard terminology and asking questions are observed and recorded in the analysis of team coordination.

Brannick et al. (1995) evaluated a method similar to the ACOE using multitrait-multimethod validation strategies. They analyzed external judges' ratings of aircrew performance along the six dimensions of

(1) assertiveness,
(2) decision-making,
(3) adaptability,
(4) situation awareness,
(5) leadership, and
(6) communication.

The aircrews flew two missions that were designed by pilots and psychologists to ensure that events in the scenario would tap into relevant teamwork behavior and to ensure that the scenarios were realistic and appropriate to the level of skill of the pilots. Judges first used behavioral observation rating forms that listed specific behaviors by skill dimension and noted when particular behaviors were observed. Judges completed these observation ratings using video recordings and were allowed to re-play segments of video as needed. After the behavioral ratings of the missions were complete, the judges studied a rating guide that linked behaviors to a numerical evaluation for each of the teamwork dimensions.

The results of the analysis showed good convergent and discriminant validity between judges. They also showed that some process measures discriminated between student and instructor teams. However, the results also showed poor convergent validity between the two scenarios flown. The authors concluded that this set of team process behaviors can be rated validly but that multiple observations are necessary to assess characteristics of individual teams with any accuracy.

3.3.3. Workload

Measures of operator workload are commonly used in studies of human-machine systems. A number of measures including primary task performance, secondary task performance, physiological measures, and subjective workload ratings have been validated (Lysaght, et al., 1989; Wickens, 1992). The measurement of team workload, as opposed to measurement of individual workload, is relatively rare (Bowers, Braun, and Morgan, 1997). Bowers, Braun, and Morgan (1997) state that team workload is the combination of two broad categories of workload – *taskwork* and *teamwork*. Taskwork refers to the workload associated with the individual tasks each team member must perform. Teamwork refers to "the interpersonal interactions among individuals that are necessary for exchanging information, developing and maintaining communication patterns, coordinating actions, maintaining social order, and so on" (Bowers, Braun, and Morgan, 1997).

Bowers, Braun, and Morgan (1997) reviewed team research that attempted to measure *team workload* as a construct, rather than just considering workload as assessed separately by individuals on the team. In one study (Thornton et al., 1992), researchers collected individual assessments of workload as well as asked individual team members to rate, using the NASA-TLX rating scale (Hart and Staveland, 1988), their perception of overall team workload. Thornton et al. evaluated several representations of workload including: (1) TLX score of highest performing member, (2) TLX score of lowest performing member, and (3) average TLX score. They found no significant correlation between task performance and workload ratings and found no significant difference in workload ratings for two manipulations of demand (low and high) for the task.

In another study (Urban et al., 1995), team workload was assessed through team member ratings using the NASA-TLX and through a secondary monitoring task. While the TLX rating scores for the team (lowest workload, highest workload, and average workload) did correlate with performance, workload measures failed to yield significant differences between low and high task demand conditions. Bowers, Braun, and Morgan (1997) concluded that assessments that rely on averaging team member assessments of workload are likely to reduce the variance such that effects of task manipulation are masked. They suggest that exploration of additional measures of workload derived directly from models of team workload is needed. They also propose the creation of new instruments that assess both coordination demands and task demands and create a composite index of team workload.

One possibility would be to simply ask individuals to rate separately the workload associated with *taskwork* and the workload associated with *teamwork*. Clay (1994) showed that individuals are capable of assessing the workload associated with two different components of a task in subjective ratings. In addition, the workload for individual team roles could be evaluated separately, rather than attempting to combine them into a single measure. One would expect that the *taskwork* workload ratings could be very different for different roles, while there should be some correlation between roles in *teamwork* ratings.

3.4. Characteristics of High Performing Teams

Researchers have evaluated the influences of task, work, individual, and team characteristics on team processes and outcomes. The results of these studies have allowed researchers to identify a number of characteristics that are associated with high performing teams.

3.4.1. Communication Behaviors

Team communication is a critical factor in team performance. As mentioned previously, a large proportion of human error is attributed to problems with team communication (Billings and Reynard, 1984; Helmreich and Foushee, 1993). Kanki and Palmer (1993) describe five functions of communication:

(1) provides information,
(2) establishes interpersonal relationships,
(3) establishes predictable behavior patterns (standard operating procedures, conventional speech),
(4) maintains attention to the task and monitoring, and

(5) is a management tool.

Studies of team communication and coordination have noted specific types of communication and coordination behaviors that are associated with good team performance (Costley, Johnson, and Lawson, 1989; Foushee et al., 1986; Orasanu, 1990). Quantity of communication has received mixed results in terms of its relationship with team performance (Orasanu, 1990; Wiener, 1993). However, one effect of quantity of communication that has been consistent across different studies is that with increased workloads, higher performing teams make fewer verbalizations than lower performing teams (Urban et al., 1993; Wiener, 1993). Related to this, Orasanu (1990) found that low performing teams increased communication during high workload, but the communication was less effective than that of high performing teams.

Researchers have found that teams in which members provide unsolicited information to other team members generally perform better than those that do not (Johannesen, Cook, and Woods, 1994; Urban et al., 1993). Johannesen, Cook, and Woods reported that in addition to providing unsolicited reports, experienced team members provided additional information in their reports. When asked a question, rather than simply answer the question, they often provide additional details based on the context of the question. Urban et al. found that good teams appeared to be more efficient in their use of questions, asking fewer questions (yet still receiving all the necessary information). Jentsch et al. (1995) found that teams that were faster in detecting a problem used more standard communications, made more leadership statements, and vocalized more situation awareness observations than did slow teams.

Researchers have also found some differences in explicit versus implicit communication of teams associated with differences in performance. Serfaty, Entin, and Volpe (1993) found that teams that switched from explicit to implicit methods of communication as time pressure increased were able to maintain a consistent level of performance. The lower communication rate of high performing teams in high workload situations over low performing teams also implies that communication occurs more implicitly in those teams.

The results of these empirical studies on team communication confirm the summary of Kraiger and Wenzel (1997) that "teams perform more efficiently when their communication is well coordinated with little excess chatter and concise statements of questioning, feedback, and confirmation".

3.4.2. Planning and Situation Assessment

Situation awareness can be defined 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" (Endsley, 1988). Endsley (1995) also defines *team situation awareness* as "…the degree to which every team member possesses the SA [situation awareness] required for his or her responsibilities." Thus, team situation awareness is not dependent on how much or how little information overlap there is between team member situation awareness requirements, only that each member of the team have complete situation awareness for his or her particular role (Endsley, 1995; Kaber and Endsley, 1998). If several team members must be aware of the same information for their role, it is not sufficient that one team member have this knowledge. Conversely, if only one team member is required to know certain information, it is not necessary that other team members be aware of this information.

Endsley (2000) makes a distinction between the process of achieving situation awareness (that is, situation assessment) and the end product of situation awareness. Team research has shown that high performing teams exhibit behaviors such as situation assessment and planning that help to achieve and maintain situation awareness. Orasanu (1990) found that planning during low workload phases of a task was associated with high performing teams. The observation by Jentsch et al. (1995) that high performing teams vocalized more situation awareness observations also supports this theory. Endsley (2000) suggests that operators rely on each other for confirming their own situation awareness. One example of the failure of a team in this regard is the crash of the Air Florida flight into the Potomac in which the voice recorder indicated that the first officer was aware of a problem that he failed to make clear to the pilot (Prince and Salas, 2000).

Related to the issue of maintaining situation awareness, the management of workload between crewmembers is important for effective team performance. Ruffell-Smith (1979) found that maintaining awareness of other crewmembers workload, whether it is to avoid overloading them or to offload tasks when overloaded, was an important skill in maintaining effective team performance.

3.4.3. Shared Mental Models

In addition to individual team member skill and knowledge, the correspondence between the knowledge of individual team members affects team performance. The term *mental model* is used to describe internal models, or schemata, stored in long-term memory that represent a person's underlying knowledge about specific systems or environments. Serfaty, Entin, and Volpe (1993) found that discrepancies between the mental model of a team leader and the mental model of subordinates on the cost of errors generated non-trivial patterns of error making in teams. This research supports their premise that effective teams develop a shared or common mental model of their common task that allows them to use the team structure to maintain team performance and coordination under a wide variety of conditions.

Stout and Salas (1993) report on research by Volpe, Cannon-Bowers, Salas, and Spector in which teams with high interpositional knowledge used more effective communication than teams in which members did not have a high knowledge of other team member roles. Kanki and Palmer (1993) summarize research that has shown that crews that are familiar with one another and crews that share similar communication patterns exhibit better team performance than crews that are unfamiliar and do not share similar communication patterns. Travillian et al. (1993) and Volpe et al. (1996) have shown that cross training team members on other team member tasks has lead to improved team performance and communication. Orasanu (1990, 1993) supports the notion of shared background knowledge and shared problem models as a positive factor in effective crew performance. She found that effective flight crews had captains that exhibited more planning behavior that facilitated the development of shared mental models between crew members.

Kaber and Endsley (1998) relate the concept of *shared situation awareness* to common mental models. According to Kaber and Endsley (1998), the sharing of situation awareness on elements of mutual interest between team members constitutes much of team coordination and is very important to team functioning and coordination.

Bolstad and Endsley (1999a, 2000) have explored the effect of shared displays on fostering shared mental models and shared situation awareness. They found (Bolstad and Endsley, 1999a) that performance was best when subjects were able to view one another's displays initially and then perform without viewing one another's displays. This suggests that sharing the displays allowed them to develop similar mental models, but that during task performance, viewing both displays provided excess information that was detrimental to performance. In a separate study (Bolstad and Endsley, 2000), teams performed better under high workload when team members used displays that abstracted pertinent information from the other team member's display, compared to fully shared displays or displays that did not provide the additional information. They also found that communication shifted from verbal communication to more implicit coordination (due to shared situation awareness) with the shared displays. At high levels of workload, the shared displays provide advantages by reducing communication requirements. Bolstad and Endsley (2000) caution that shared displays should be designed carefully because excess information provided on shared displays can slow performance.

3.5. Summary

Clearly, assessing system performance within the context of teams is important given the operating environment of complex systems. A definition and model of team effectiveness has been presented. Based on this, researchers have developed a number of techniques for measuring teams. These techniques include the measurement of task outcomes, the measurement of team processes, such as communication and coordination, and the measurement of team workload.

Because of the need to understand the underlying functioning of teams, researchers agree that team process measures are important in the study of team performance (Brannick and Prince, 1997). The measurement of team communication and coordination can be accomplished through objective measures such as quantifying and classifying verbalizations and gestures. However, these types of measures often fail to illustrate the underlying causes of differences in communication rates. In addition, studies have shown contradictory results when the quantity of communication is measured (Orasanu, 1990; Wiener, 1993). Subjective measures of crew coordination through the use of external judges may provide additional insight into the underlying functioning of teams. Researchers have identified a number of constructs that appear to differentiate the performance of teams (Bowers et al., 1992; Brannick et al., 1995; Prince, et al., 1992). Some of these constructs include: (1) assertiveness, (2) decision-making, (3) adaptability, (4) situation awareness, (5) leadership, and (6) communication.

Research in the areas of team performance has identified behaviors associated with these constructs that are representative of high performing teams. First, high performing teams exhibit efficient verbal communication, verbalizing task relevant details with little irrelevant information, often without specific requests for information (Johannesen, Cook, and Woods, 1994; Urban et al., 1993). Related to this, high performing teams reduce their quantity of verbalizations as workload increases and use more standard communications (Jentsch et al., 1995; Urban et al., 1993; Wiener, 1993). This type of communication likely decreases the workload associated with teamwork, such as attending to questions and answers, while still providing needed information allowing teams to maintain a high level of performance.

Second, high performing teams conduct frequent planning and situation assessment sessions (Jentsch et al., 1995; Orasanu, 1990). Planning and situation assessment serves to assist the team in maintaining a high state of situation awareness. Finally, shared or common mental models between team members result in better team performance. This is likely due to reduced crew communication requirements because of the common knowledge. Shared models may be developed through training, via team behaviors such as planning, or via display design.

Up to this point, this dissertation has presented an overview of research in the areas of automation and teams independently. While research on the interaction of automation and teams is limited, there is some information available to indicate that there may be qualitative differences in the way teams coordinate within an automated environment (Mosier and Skitka, 1996; Wiener, 1993). The following section summarizes the current literature that evaluates teams in the context of complex automated systems.

4. AUTOMATION AND TEAMS

Several researchers have stressed the importance of evaluating the relationship between automation and teamwork (Bowers et al., 1996; Kaber et al., 2001; Wickens, Mayor, and McGee, 1997). Cannon-Bowers, Salas, and Converse (1993) suggest that automation may adversely affect team performance because automation adds an additional requirement that team members have a mental model of the automated system, in addition to their model of the task, equipment, and the team. In addition, the change in the form of information presentation generally associated with automation (e.g., a single multi-function electronic display compared to multiple mechanical displays) is likely to affect team coordination. This may result in a reduced ability to physically observe the actions of other team members (Johannesen et al., 1994; Wiener, 1989, 1993). Also, communication may be completed electronically through the automated system, rather than verbally (Kaber et al., 2001). Thus, the design of automated systems plays a role in team coordination.

Most of the research on automation and teams to date has been in the form of surveys and field or simulator evaluations of flight deck automation. These evaluations provide initial support for the notion that automation has some type of effect on team coordination. However, the results are mixed in terms of "how" automation affects teams.

First, it is important to note that the limited studies on automation and teams have shown few differences in terms of team effectiveness or outcome measures between teams in automated and non-automated conditions. Wiener et al. (1991) found no difference in team outcome measures between an automated (MD-88) and a non-automated (DC-9) aircraft. Similarly, in a simulator study, Bowers et al. (1993) found that automation was not related to more effective task performance.

4.1. Automation and Team Coordination

Research has shown significant differences in terms of crew coordination based on team process measures. Wiener (1989) found that surveyed pilots reported greater difficulty associated with automated aircraft in delegating programming and monitoring responsibilities among crew members. Most of the results related to crew coordination are based on measures of communication.

Several studies have reported on differences in the quantity of verbal communication between automated and non-automated team performance conditions. In a survey by Wise et al. (1992), corporate pilots reported an increased need for crew communication associated with operating automated aircraft. Straus and Cooper (1989) also found that verbal communications was more prevalent among flight crews in automated conditions. In contrast, Costley, Johnson, and Lawson (1990) performed a field study comparing the more advanced B757 with the traditional cockpit of the B737-200 and found that lower communication rates were associated with more automated aircraft. These contradictory results on the quantity of communication suggest that it is important to look more deeply into what types of team communications are affected by automation.

First, there appear to be differences related to characteristics of the team itself. Heterogeneous crews appear to communicate more with each other in automated conditions than do homogeneous crews (Kaber et al., 2001; Petridis et al., 1985; Straus and Cooper, 1989). In these studies, it appears as though the more experienced member on the team verbalizes more in terms of explanations to the novice member. Bowers et al. (1993) compared the performance and communication of crews in simulated flights with and without automation. In manual aircraft, there was little difference in communication patterns between poor performers and good performers. However, in the automated condition poor performers made more unsolicited observations and responses to these observations than did effective performers.

There also appear to be conflicting results with respect to the differences in types of communication observed in automated versus manual aircraft. In the Costley, Johnson and Lawson (1990) study, they found that questioning appeared to be the primary difference in communication rates, with fewer questions associated with automated aircraft. Veinott and Irwin (1993), in a further analysis of the Wiener et al. (1991) comparison of the MD-88 and DC-9, found that crew members in the conventional aircraft (DC-9) relied more on command-acknowledgment sequences than did crew members in the advanced cockpit.

Clothier (1991) analyzed data from line oriented flight training in simulators and in airline operations comparing advanced and conventional cockpits. Clothier found that, in simulations, automated cockpits were associated with higher quality communication as shown by ratings of crew coordination skills such as inquiries, advocacy, and avoiding or prioritizing distractions.

These results appear to provide conflicting evidence not only on the quantity of communication but also on the quality. Some studies show automated aircraft increasing communication, which may be reflective of poor design or limited understanding of the automated system (Bowers et al., 1993; Petridis et al., 1985; Straus and Cooper, 1989) and other studies show decreases in communication or better quality of communication associated with automation (Clothier, 1991; Costley, Johnson, and Lawson, 1990; Wiener et al., 1991). This raises the question of whether differences between the aircraft other than automation (such as advanced displays, or the type and combination of automated systems) may have been associated with the communication differences (Costley, Johnson, and Lawson, 1990). Wiener (1993) concluded from the MD-88 and DC-9 study and from the study by Costley, Johnson, and Lawson that automated systems may have the unintended consequence of interfering with team communication and coordination. He stated that flight deck equipment and configuration affect the quality and perhaps quantity of communication and crew coordination in ways that are not yet well understood.

Another potential reason for differences in quantity and type of verbal communication may be due to changes in non-verbal communication with automated systems. Non-verbal communication between team members such as interacting with displays and controls provides important information to team members. According to Bowers et al. (1996) and Wiener (1993), automation interferes with non-verbal communication because it is physically difficult for team members to see what each other is doing (Bowers, et al., 1996; Wiener, 1993). Kaber et al. (2001) also concluded that teams using automation are apt to use the automated interfaces as a means of communication as opposed to communicating verbally.

One consistent result that can be found across studies on verbal communication and automation is that there is a clear interaction with workload. Significant differences between systems tend to appear when workload is higher. For example, Clothier's (1991) results held up only in the high workload simulation and did not appear in actual flight. In addition, Costley, Johnson, and Lawson (1990) also compared the B737-300 and the B737-200 and found that an increase in communication during climb and descent (high workload phases) due to the automated system, but a slightly lower rate of communication during the cruise phase.

Table 4 summarizes the results of these and a few other research studies of automation and team communication and performance.

Table 4. Summary of automation and team communication/performance research.

Reference	**Method**	**Team Performance Results**	**Team Communication Results**
Costley, Johnson, and Lawson, 1989	Aviation field study		Trend toward lower communication rates with more automated aircraft.
Orlady, 1989	Pilot survey		40% of pilots report substantial differences in CRM activities associated with automation, 40% report no difference, 20% no opinion.
Clothier, 1991	Aviation simulator and field study		High quality communication associated with automation in simulated (high workload) flights but not in actual flights.
James et al., 1991	Pilot survey		Minor indication that more pilots thought automation increases the quality of crew communication.
Wiener et al., 1991 and Veinott and Irwin, 1993	Aviation simulator study	No difference between automated and conventional aircraft	Increased crew communications and questions associated with automated system.
Wise et al., 1992	Survey of corporate pilots		Increased need for crew coordination.
Bowers et al., 1993	Aviation simulator study	No difference between automated and non-automated conditions	Poor performers in the automated condition made more unsolicited observations and related responses.

Bowers et al. (1996) concluded from his review of team performance in automated systems that "the current knowledge base regarding automation and team performance is woefully inadequate". Bowers et al. suggested a few preliminary hypotheses based on the available data. Three of their six hypotheses were:

1. Teams working with automatic systems might benefit from briefings designed to articulate team norms.
2. Teams in automatic systems will have to communicate differently to maintain effective performance.
3. The effects of automation on team performance will be most salient in high workload conditions.

4.2. Summary

This section reviewed literature that has assessed the affect of automation on team performance and team coordination. The research indicates no conclusive findings of the effect of automation on team outcome measures (Bowers et al., 1993; Wiener et al., 1991). The results are also mixed in terms of quantity and quality of communication. Automation has been seen to both increase and decrease the amount of communication (Costley, Johnson, and Lawson, 1989; Wiener et al., 1991; Wise et al. 1992) and automation has been associated with both improvement and worsening of the quality of communication (Bowers et al., 1993; Clothier, 1991; Costley, Johnson, and Lawson, 1989; Petridis et al., 1985; Straus and Cooper, 1989; Wiener et al., 1991).

The varied results of these studies raise a number of questions related to automation and teamwork. Are the differences in these results due to fundamental characteristics of automation or due to the effects of programming requirements or differences in the task? Do the effects seen in these studies generalize to automated systems other than flight deck automation? How do non-verbal methods of communication affect team performance and how does the presence of automation in complex systems interact with these methods of communication?

5. PROBLEM STATEMENT

Studies of the effects of automation on individual task performance have presented a number of negative consequences associated with automated control systems (Bainbridge, 1987; Funk, Lyall, and Niemczyk, 1997; Woods, 1996). By studying these problems, researchers have also been able to begin to identify potential solutions (Bainbridge, 1987; Bowers, 1996; Endsley and Kaber, 1999; Kaber, 1997; Scerbo, 1996; Wiener, 1993; Woods, 1996). Because automation affects individual task performance, it is also expected that automation may have some effect on the performance of teams. In fact, it is possible that human performance consequences associated with automation may be more extreme in team environments because of the added complexity due to task coordination.

A large percentage of human error in team tasks is attributable to communication between team members (Billings and Reynard, 1984; Helmreich and Foushee, 1993); therefore, it is important to ascertain the effect of automation on team communications. A number of researchers have speculated that automation may have an adverse effect on team communication, either due to increased requirements for crew coordination due to interpreting the automation (Cannon-Bowers, Salas, and Converse, 1993; Wiener, 1993) or due to changes in the nature of the communication because of reduced ability to physically observe the actions of other team members (Johannesen et al., 1994; Wiener, 1989, 1993).

Research on teams in a number of different environments has identified characteristics of high performing teams. These characteristics include: (1) effective communication with little extraneous information; (2) frequent situation assessment and planning; and (3) shared or common mental models between team members. Empirical research on the interaction of automation and teams has yet to determine the role of automation in these characteristics of high performing teams.

Some work has evaluated the effects of automation on team performance and team coordination, but the experiments have been limited to the environment of advanced aviation (Bowers et al., 1993; Clothier, 1991; Costley, Johnson, and Lawson, 1989; Wiener et al., 1991). Most of the research that is available on the aviation domain compares teamwork in early model, conventional aircraft with the newer advanced models of the same aircraft. Differences between the two models are often numerous and go beyond the addition of more automation (e.g., changes in functionality, displays, and controls). Therefore, it is difficult to determine what the true associations are between team communication and performance and specific characteristics of the aircraft, including automation. The conflicting results of the studies to date (Bowers et al., 1993; Clothier, 1991; Costley, Johnson, and Lawson, 1989; Petridis et al., 1985; Straus and Cooper, 1989; Wiener et al., 1991) support the notion that we have not yet isolated team performance and team coordination effects due to automation.

In order to isolate the effects of automation on human performance, we must first establish a clear definition of automation. Researchers (Endsley and Kaber, 1999; Parasuraman et al., 2000; Scerbo, 1996; Wickens et al. 1998) have begun to reach a consensus on a model of types and levels of automation (LOA) that can be used to classify modes of system automation from low to high across human-machine information processing functions including: (1) information acquisition, (2) information analysis, (3) decision and action selection; and (4) action implementation. This model has been used in the study of the effects of LOA on individual task performance (Endsley and Kaber, 1999; Horrey and Wickens, 2001; Kaber and Endsley, 1997a, 1997b). There have as yet been no studies of the effects of automation on team performance that use this theoretical model as a basis for defining automation conditions.

The present study evaluates the effect of automation on team coordination and performance in a decision-making task. In order to more completely describe the effects of automation, the task was designed to present different levels of automation including computerization of: (1) information acquisition; (2) information analysis; and (3) decision action and selection functions (according to the model of types and levels of automation presented by Parasuraman et al. (2000)). The automation conditions presented can be roughly classified into categories based on the theoretical taxonomy of levels of automation presented by Endsley and Kaber (1999) (see Table 2).

The objective of this study was to assess the effects of different types and levels of automation on the performance and coordination of teams, and based on this analysis, to identify ways in which human-automation interaction can be improved in a team environment. Questions to be addressed by this study include whether team members use different communication patterns with and without automation and to what extent do different LOAs affect characteristics known to be important to good teamwork.

It was expected that increasing automation of information acquisition, information analysis, and/or decision selection would lead to increased team effectiveness. As far as team coordination, automation was expected to qualitatively change the way in which team members communicate. Differences between the different automation conditions were expected for team processes associated with situation assessment, decision-making, leadership, and communication. In accordance with results found for individual performance (Kaber et al., 1999), team member situation awareness was expected to be higher in the automation of information acquisition and analysis (intermediate forms of automation) than in the automation of decision-making. Consequently, higher ratings of decision-making and leadership were expected for information acquisition and information analysis compared to other conditions. Because automation of information acquisition helps to support shared awareness between team members, verbalization of situation assessment was expected to be lower for this condition. The information acquisition condition was also expected to result in greater effectiveness of communication (e.g., more task relevant statements and fewer irrelevant or extraneous statements) compared to automation of information analysis and decision-making. In addition, the shared information provided to both team members in the information acquisition condition was expected to support coordination between the team members, resulting in better ratings of decision-making and leadership.

6. METHOD

6.1. Task

The team performance task is based on a laboratory simulation of a Theatre Defense Task (TDT). The TDT was developed by Bolstad and Endsley (1999a), and is based on an individual control task developed by Kaber and Endsley (1997). A number of modifications to the original TDT by Bolstad and Endsley (1999a) have been made for the purposes of this experiment. The TDT is a team decision-making and target elimination task completed by two team members. The team members, an Intelligence Officer (IO) and an Air Commander (AC), have separate, but inter-related tasks. The team members work at separate workstations, connected by an Ethernet-based Local Area Network. The role of the AC is to protect the home base from incoming aircraft. The role of the IO is to classify incoming targets as enemy or friendly and to indicate the type of aircraft. Based on the aircraft type, the AC must choose an appropriate missile to destroy enemy aircraft, or allow friendly aircraft to pass through.

The task is relatively complex requiring the IO to consider several data sources simultaneously, and account for the sensitivity of the sources, in determining the classification of a target. The task requires communication between the AC and the IO. Some communication is via the AC and IO computer interfaces with the AC sending requests for target classifications to the IO's display and the IO sending classification information to the AC's display. In addition, verbal communication is required from the AC to the IO regarding target positions and airborne warning and control system (AWACS) reconnaissance aircraft positions due to differences in sensor reliabilities based on the positions of sensors. Verbal communication may also be used for setting priorities and correcting errors such as misclassification of targets. The pace of the task is such that both team members must be strategic about how they prioritize and process targets in order to achieve a high score. Otherwise, a significant number of enemy aircraft will penetrate the air defense and strike the home base, or a significant number of friendly aircraft will be destroyed.

Figure 2 presents the AC's task display. The display includes a radarscope of the surrounding 30 miles of airspace, with the scale displaying "tens" of miles. Each square on the display represents a target. Targets can be classified as fighters, transports, or bombers and can be either enemy craft or friendly. Targets first appear on the AC's display as white squares. The AC may request a classification of a target verbally, or by entering a target number using the number keypad and then pressing the enter key. If the AC makes a request through the interface, the request is sent to the IO's display. Once the IO has classified a target based on the available sensor information, he or she may inform the AC verbally or may use the interface to send the classification information to the AC. If the IO uses the interface, the target color changes based on the classification. Table 5 indicates the colors for the various target classifications.

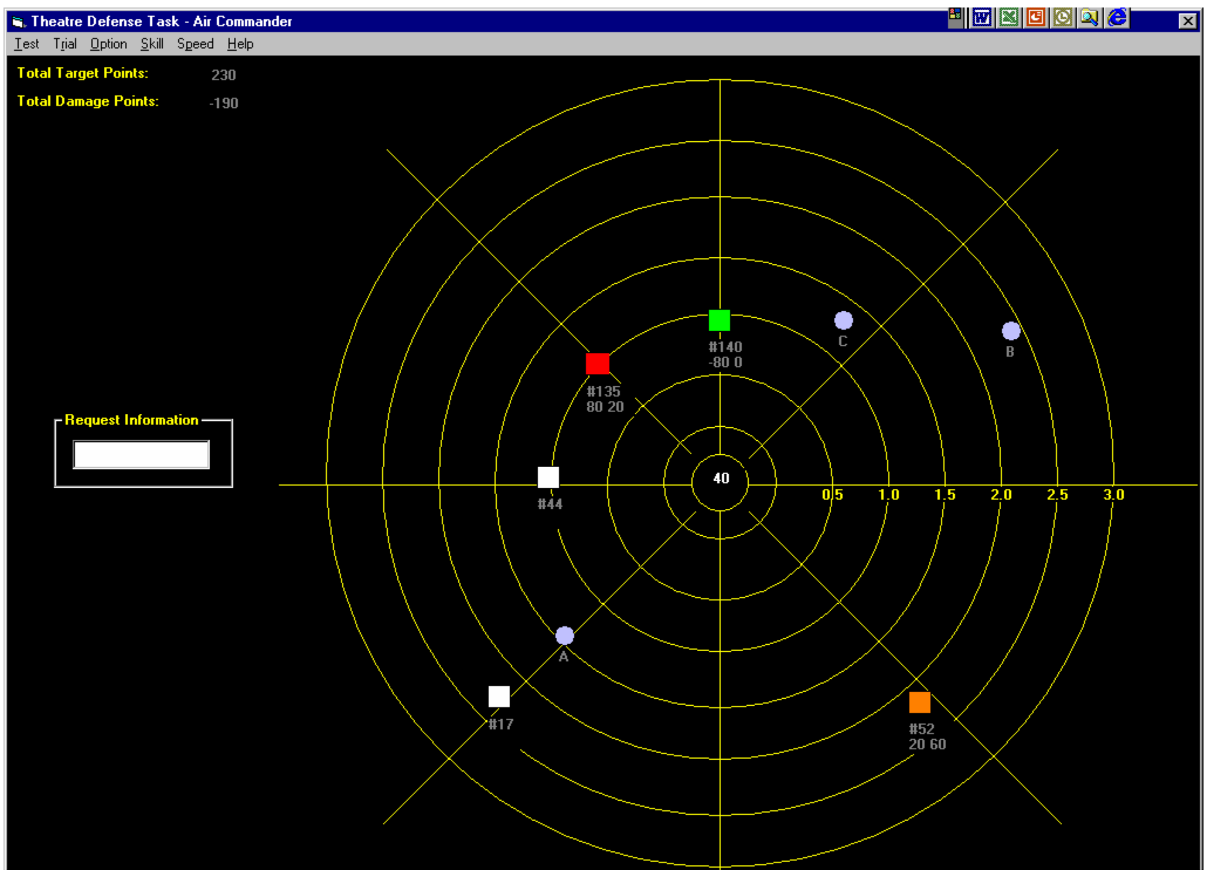

Figure 2. Air Commander workstation's screen.

Table 5. Target classification details.

Plane Categories	**Types**	**Color**	**Target Points**	**Damage Points**
Friendly Fighters	F/A-18	Blue	-20	-
	F-15E	Blue	-40	-
	F-16	Blue	-60	-
Friendly Bombers	B-52	Green	-50	-
	B-1	Green	-80	-
	B-2	Green	-100	-
Friendly Transports	C-130J	Turquoise	-120	-
	C-21	Turquoise	-140	-
	KC-135	Turquoise	-150	-
Enemy Fighters	Mig-29	Red	60	10
	Su-35	Red	80	20
	Su-37	Red	100	10
Enemy Bombers	Tu-22M	Orange	10	50
	Tu-168	Orange	20	60
Enemy Transports	An-124	Yellow	50	50
	An-225	Yellow	60	60

The AC has two types of simulated missiles available: a Sparrow which is smaller and faster and used for destroying fighters and an AMRAAM (armed medium range air to air missile) which is a long range radar guided missile used for destroying bombers and transports. The left mouse button launches Sparrows and the right mouse button launches AMRAAMs. If the appropriate missile is launched, the plane is destroyed. If an incorrect missile is launched for the targeted aircraft, there is a 50% chance that the plane will be destroyed.

In addition to the targets on the radarscope, the position of three AWACS reconnaissance aircraft is displayed (designated as A, B, and C in Figure 2). The AWACS aircraft provide plane identifications and relay this information to the IO. The reliability of the information is dependent upon the location of the AWACS aircraft. Aircraft that are closer to the home base provide more reliable information. From 0 to 10 miles away from the base, the reliability is 90%; from 10 to 20 miles away, the reliability is 60%; and from 20 to 30 miles away, the reliability is 30%. The reliability of the AWACS sensors is based only on the distance from the center and is in no way related to the proximity of the sensor to particular targets. The AWACS aircraft initially appear at random positions and their movements on the radar screen over time are randomly generated.

The final information provided on the AC's display is feedback regarding target points and damage points. Damage points are given when an enemy aircraft reaches the home base. Target points are given when an aircraft is destroyed. Positive target points (rewards) are given for destroying enemy aircraft, while negative target points (penalties) are given for destroying friendly aircraft. The number of points given is dependent on the mission relevance and the lethality of the aircraft (see Table 5). For example, a large penalty is assessed against target points for destroying a friendly transport and a heavy damage is assessed for allowing an enemy bomber to reach the home base.

Figure 3 presents the IO's task display. Along the top of the display are boxes for listing target classifications requested by the AC. In the upper left hand corner of the display is the information provided by the AWACS aircraft regarding target classification. To view the AWACS information for a particular target, the IO must select the button to the left of the target number of interest. The sensor information provided includes a classification by the three AWACS sensors, depending on the position of the sensors as described previously. For example, if a sensor's reliability is 60%, then there is a 60% chance that the aircraft classification reported under that sensor will be correct (and there is a 40% chance that the sensor will randomly display any of the other 16 aircraft). The information provided on the IO display is based on the location of the sensors at the time the target first appeared on the screen and that information is maintained for the duration of the time that the target travels toward the home base, independent of any movement of the AWACS sensors.

The random movement of the AWACS sensors results in an average reliability for a single sensor over a 15-minute trial of approximately 48%. Thus, on average, over a 15-minute trial, the chance of all 3 sensors reporting correctly is about 11%; the chance of 2 sensors reporting correctly is about 36%; the chance of 1 sensor reporting correctly is about 39%; and the chance that no sensors will report correctly is about 14%. In the case where only one sensor reports correctly, the IO still has a good chance of guessing correctly, since he or she knows which of the three sensors is most reliable (if the information is provided correctly by the AC).

The interface for classifying a particular target is located in the upper right hand section of the IO's display. After the IO has selected a particular aircraft in the left hand section of the display, reviewed the available sensor information, and made a decision regarding the classification; he or she then presses the appropriate button to classify the aircraft. This allows the classification information to be passed to the AC's display in the form of color-coding.

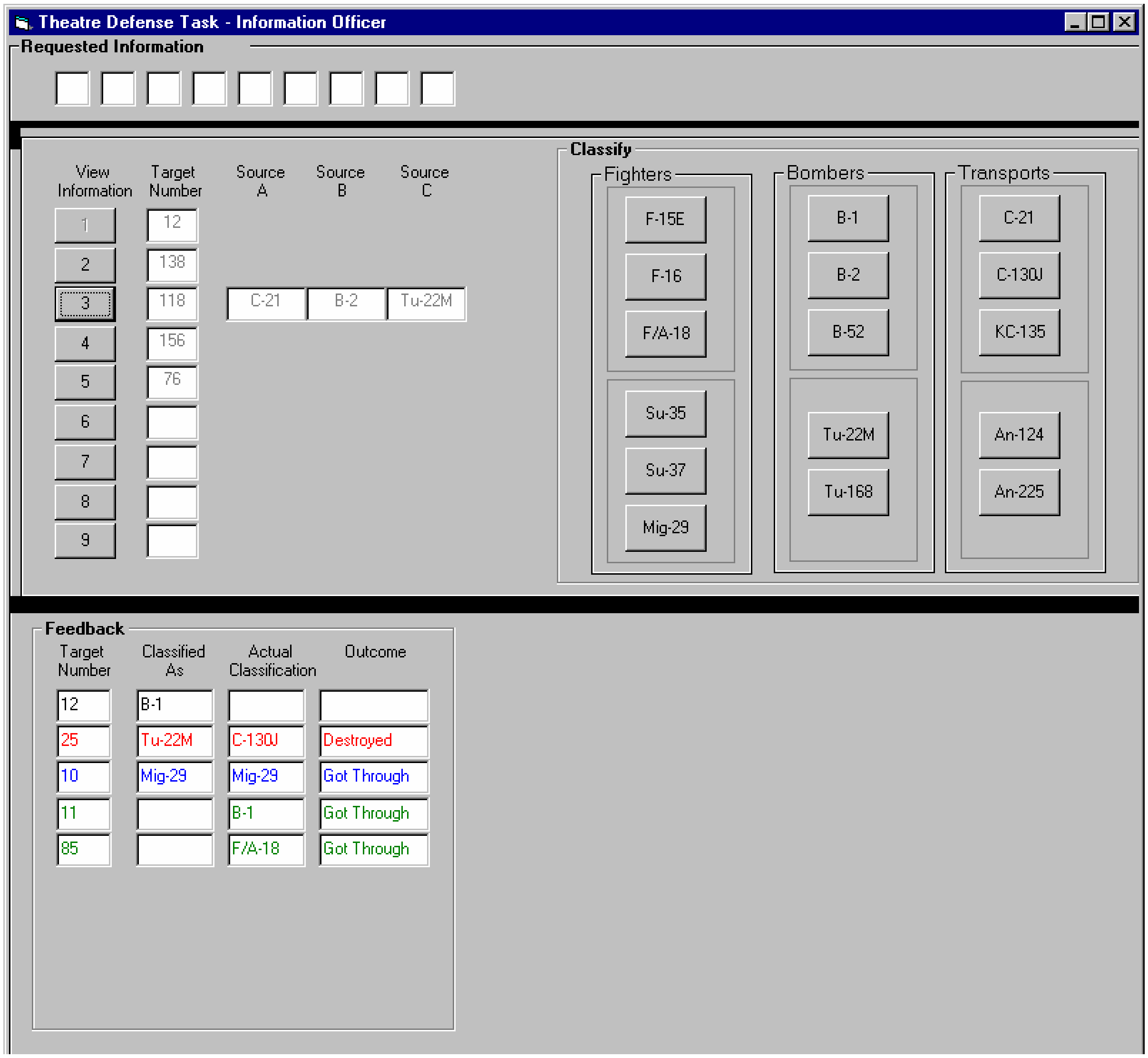

Figure 3. Intelligence Officer's workstation screen.

The lower section of the display provides feedback to the IO with respect to the classifications that have been made and the outcome of the target. This feedback includes the classification determined by the IO, the actual classification of the target, and the outcome based on the AC's response to the target. The feedback text is color coded to indicate the success of the IO's classification. Correctly classified targets are colored blue, incorrectly classified targets are colored red, and targets that were not classified are colored green (see Figure 3). The colors indicate only the success of the IO's classifications and do not reflect whether the final outcomes (based on the actions of the AC) were correct.

A number of factors must be considered in the decisions of both the AC and the IO. The IO must decide on the order in which to classify targets. He or she may base this decision on requests from the AC displayed on his or her screen and on verbal communication from the AC regarding target positions. In addition, the IO may also review the sensor information for the targets and make decisions related to which targets to classify first based on the amount of information available for each target and the content of the information. For example, the IO may quickly scan through the target information and classify the targets with two or more sensors matching first.

Once the IO has decided to focus on a particular target for classification, he or she must consider several factors in classifying the target. The IO must consider the sensor information provided and, particularly in the case of dissonant information, must also consider the reliability of the sensors. In order to assess the reliability of the sensors, the IO must communicate with the AC.

The AC also has a number of factors to consider in completing the task. The AC must decide on the order in which to process targets. Factors that affect this decision include the position of the target relative to the home base, whether the target is involved in an impending collision, and confidence in the targets classification. As the information shown on the AC's display does not include specific sensor reports, the AC may choose to communicate with the IO regarding confidence in a specific classification. In addition, the AC must decide on which missile to use based on the color-coding of the target.

6.1.1. Levels of Automation

The TDT was modified to present four unique levels of automation for comparison of team performance and team coordination across automation types and levels. For a team task, it may be possible that one team member is presented with one level of automation and another team member is presented with a different level. For the purposes of this experiment, the automation is classified on the basis of the combined team task, rather than individually for each team member. The conditions have been defined such that the level of automation is similar for each team member in each condition.

The levels of automation to be studied for this experiment are defined in terms of the model of types and levels of automation proposed by Parasuraman et al. (2000). As discussed previously, this model suggests that automation can be classified from low to high along the four functions of (1) information acquisition, (2) information analysis, (3) decision selection, and (4) action implementation. The functions in the Parasuraman et al. model are essentially the same as those presented by Endsley and Kaber (1999), with modified terminology. The automation conditions selected for this experiment represent four unique combinations of varying degrees of automation along the roles of the first three of the four functions (information acquisition, information analysis, and decision selection). Changes in the degree of automation along these three functions were selected since they are expected to yield the greatest differences in team coordination.

The four levels of automation in this experiment can be roughly described using the taxonomy provided by Endsley and Kaber (1999) and include: (1) Action Support (AS); (2) Shared Control (SHC); (3) Decision Support (DS); and (4) Blended Decision Making (BDM). The definitions provided by Endsley and Kaber for the four levels of automation are provided in Table 6. However, the automation conditions for the present study do not fit perfectly into Endsley and Kaber's taxonomy because the "action implementation" aspect of the task does not vary between the different levels of automation proposed for investigation. For Endsley and Kaber, action implementation is shared between the human and computer in the AS and SHC conditions while action implementation is completely controlled by the computer for DS and BDM conditions (see Table 2 on Page 6). For the purposes of this experiment, action implementation is shared between the human and the computer in the same manner in all conditions. The other three functions of the task (monitoring, generating, and selecting) follow the human/computer roles as assigned for those conditions within Endsley and Kaber's taxonomy. Although the mapping of automation is not exact, the terminology developed by Endsley and Kaber will be used in this experiment so that comparisons can be made between similar automation conditions across experiments and to limit the proliferation of new terminology to define essentially similar forms of automation.

Table 6. Definitions of automation conditions from Endsley and Kaber (1999).

Action support	The system assists the operator with performance of the selected action, although some human control actions are required. A teleoperation system involving manipulator slaving based on human master input is a common example.
Shared control	Both the human and the computer generate possible decision options. The human still retains full control over the selection of which option to implement; however, carrying out the actions is shared between the human and the system.
Decision support	The computer generates a list of decision options that the human can select from or the operator may generate his or her own options. Once the human has selected an option, it is turned over to the computer to implement. This level is representative of many expert systems or decision support systems that provide option guidance, which the human operator may use or ignore in performing a task. This level is indicative of a decision support system that is capable of also carrying out tasks, while the previous level (shared control) is indicative of one that is not.
Blended decision making	The computer generates a list of decision options that it selects from and carries out if the human consents. The human may approve of the computer's selected option or select one from among those generated by the computer or the operator. The computer will then carry out the selected action. This level represents a higher level decision support system that is capable of selecting among alternatives as well as implementing the [selected] option.

The different degrees of automation for each of the four conditions are shown in Figure 4. Note that the presentation of these conditions is adapted from the Parasuraman et al. (2000) model of types and levels of automation and is intended to indicate *relative* differences in automation between conditions rather than any absolute judgment of degree of automation. The AS condition represents the lowest level of automation across all four functions. SHC represents an increase in both information acquisition and information analysis over AS. The DS condition represents an increase in automation of information analysis over both the AS and the SHC conditions, while it also represents a return to the lower level of automation of information acquisition. The BDM condition is similar to DS in automation of information acquisition and information analysis; however, BDM represents an increase in decision selection over all of the other conditions.

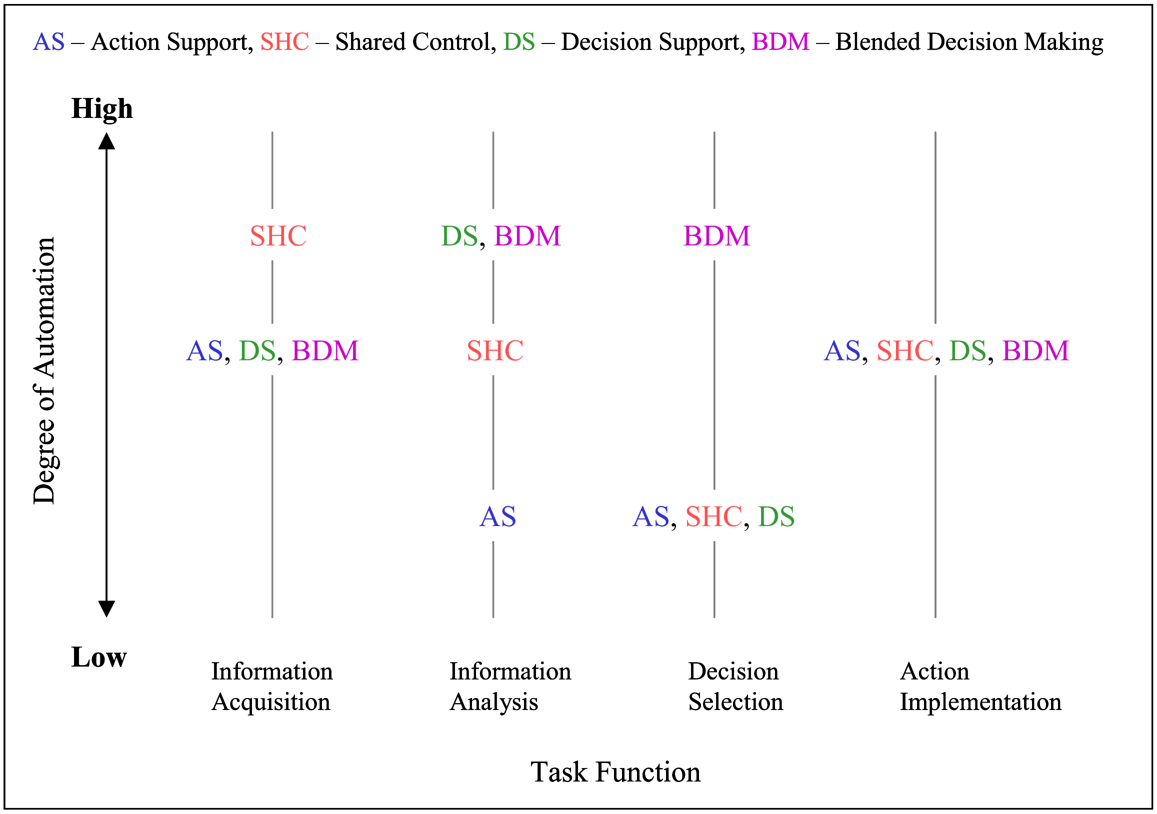

Figure 4. Mapping of automation conditions to a model of types and levels of automation.

This selection of experimental conditions allows for a number of comparisons with respect to the effect of automation on teamwork. Comparisons can be made between conditions to look at increasing automation of (1) information analysis, (2) both information acquisition and information analysis, (3) decision selection, and (4) both decision selection and information analysis. Table 7 presents the condition comparisons and the automation effects that may be evaluated.

Table 7. Level of automation comparisons.

Comparison	Increasing automation of:
AS vs. SHC	Information acquisition and information analysis
AS vs. DS	Information analysis
AS vs. BDM	Information analysis and decision selection
DS vs. BDM	Decision selection

6.1.1.1. Action Support

The AS condition represents the TDT as presented by Bolstad and Endsley (1999) with minor user interface changes. The displays and controls of the AC and IO are as described in the previous section. Targets are listed on the IO's display in the order that they enter the control area. New target information is presented on the display in the position of expiring targets. There is no automated assistance presented through the displays for either the AC or the IO in terms of prioritizing targets.

The information acquisition function of this task is shared between the human and the computer. Automation of information acquisition is provided in the form of the radar screen and target presentation for the AC. For the IO, targets are listed on the display and the sensor classification of targets is presented, when requested. However, both team members are required to communicate with one another to receive additional information in order to perform their task. The AC must receive target classification information from the IO and the IO must receive information regarding target and sensor proximity to the home base from the AC. Also, the IO must sequentially select targets of interest to display the sensor information, rather than the sensor information for all targets being automatically displayed, simultaneously.

The information analysis function of the task is primarily allocated to the human in this condition. Although much information is provided to the team members, the information is not automatically analyzed and presented in a format to aid option selection and decision-making. There is some automation of information analysis, for example, the sensor data presents the actual aircraft identified rather than source data, such as size, speed, etc. However, in general, the target displays on the IO's screen are not presented in any order (i.e., sorted by proximity from the home base) and the AC's display contains no sensor data about the targets (as is available to the IO).

The decision selection function of the task is allocated to the human operators. In the original TDT, there is no automation that classifies or prioritizes targets for either the IO or the AC; the team members must make all decisions.

Finally, the action implementation function of the task is shared between the human operators and the computer. Once the IO makes a decision regarding classification of a target, the IO sends this information to the AC by making a simple selection and button press. Update of the AC's display is automated. A similar situation exists for the AC – eliminating targets requires human input in the form of a correct mouse click, but actions such as loading, aiming, and firing missiles are all considered to be under control of the computer.

6.1.1.2. Shared Control

In a study by Bolstad and Endsley (1999b, 2000), a second version of the TDT was developed to examine the effect of shared displays on team communication. Bolstad and Endsley compared performance on the original TDT under three conditions: (1) the AC and IO could not see each other's display; (2) the AC and IO could see each other's display; and (3) the displays were modified such that the relevant information for the IO was abstracted from the AC's display and presented on the IO's display and vice versa. Bolstad and Endsley labeled this third condition of the TDT as the Abstracted Shared Display.

The Abstracted Shared Display condition designed by Bolstad and Endsley represents automation of the information acquisition and information analysis components of both the AC's and IO's task. For the purposes of this experiment, Bolstad and Endsley's Abstracted Shared Display condition, with some minor user interface modifications, was used as the SHC condition. The modified displays are presented in Figures 5 and 6. In this condition, the decision selection and action implementation portions of the task remain as defined in the AS condition. However both the information acquisition and the information analysis functions of the task are automated at higher levels.

With respect to information acquisition, all of the information necessary for the AC and the IO to perform their duties is presented on their respective displays. This eliminates the need for them to communicate, except in the case of impending collisions, planning or strategizing, and handling errors. There is also an increase in the automation of information analysis associated with the additional displays. The IO's display includes a listing of the targets based on their proximity to the home base. This gives the IO a potential priority order to follow for classifying targets. The AC's display presents information regarding sensor classification of the targets, once the IO has selected a classification. This allows the AC to double-check the IO's decisions if necessary. The additional automation of information acquisition, however, does not preclude communication between team members. Because the IO's task becomes visually intense with the additional automation information, the team members may choose to use verbal reports from the AC rather than, or in addition to, the abstracted information.

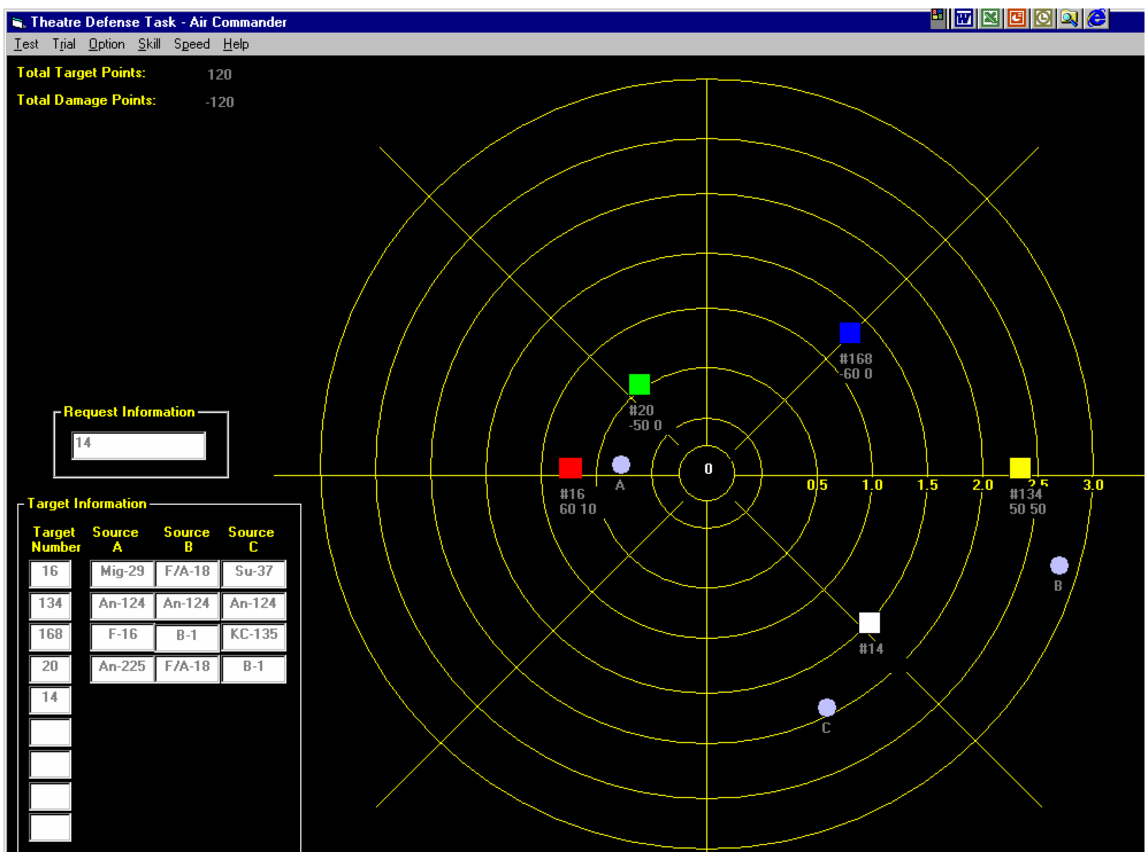

Figure 5. Air Commander workstation's screen – shared control display.

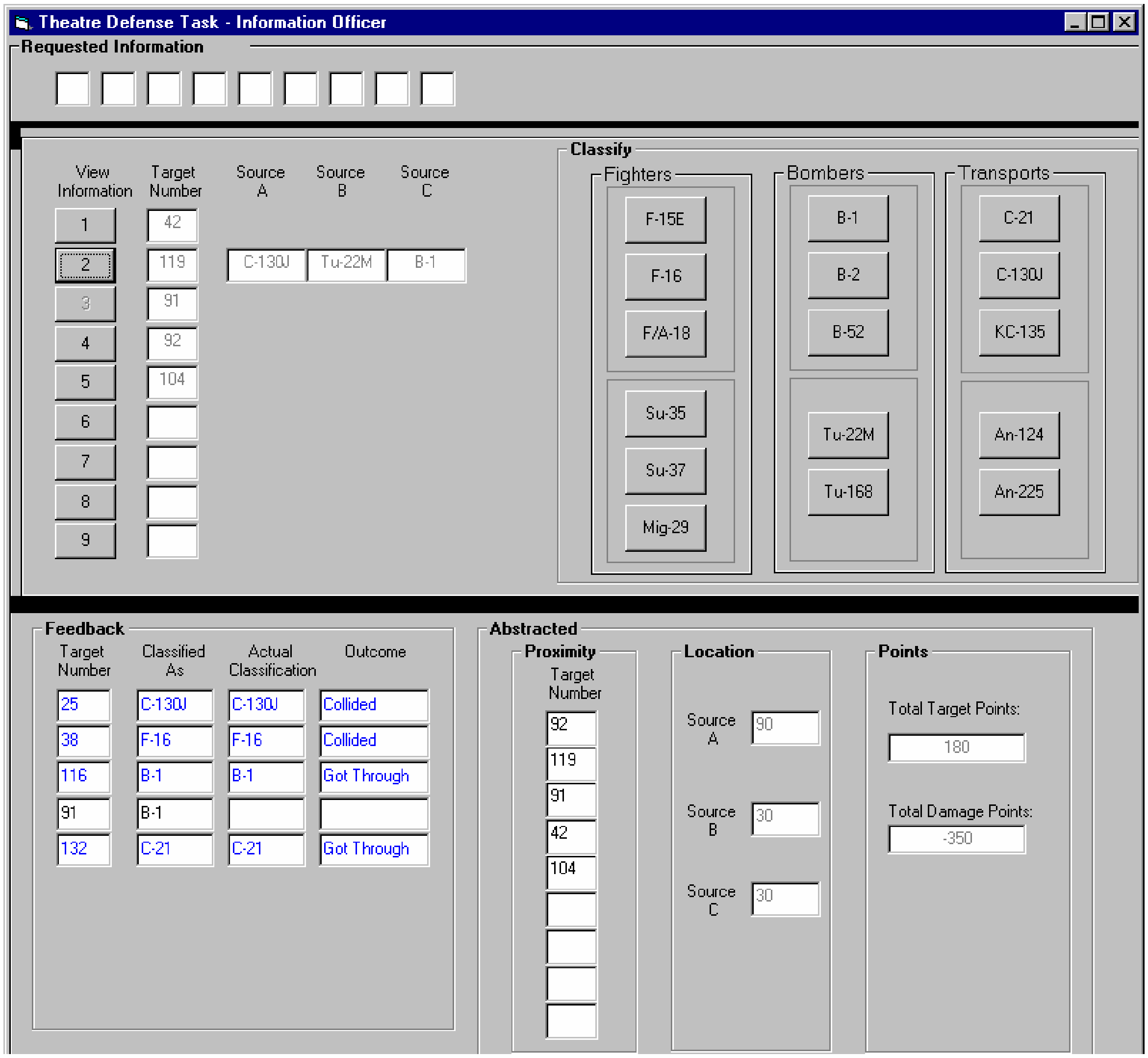

Figure 6. Intelligence Officer's workstation screen – shared control display.

6.1.1.3. Decision Support

As a third automation condition, the TDT was modified to present the IO's target list in an order indicating which target should be classified first based on its distance to the home base and target type information provided by the sensors (e.g., enemy targets close to the home base will be presented first in the list). In this case, the information presented to both the AC and the IO is the same as in the AS condition except that on the IO's display, the targets appear in a prioritized order that represents an advanced level of automated information analysis. The algorithm to determine the priority listing first assesses what type of aircraft a target is likely to be on the basis of the highest reliability sensor reading (or a combination of readings if two sensors report the same aircraft). If the target classification by the automation is “enemy”, the algorithm sorts those targets to the top of the list, with the enemy target closest to the center of the AC’s display nearest to the top of the IO’s target list. All aircraft that are classified as “friendly” by the automation will be listed after the potential enemies. The automation does not present any of the logic or details of this algorithm to the operator (e.g., what type of aircraft the automation classified the target as or what the exact position of the target is). The automation only uses this information to present a suggested order in which the targets should be processed. Because the automation sorts the list in priority order, the list is continually re-sorted every time a target reaches the center of the display or is eliminated. This can cause increased time pressure for the IO since the information displayed on a particular row of his or her screen may disappear when the list is re-sorted (and he or she would have to find the target in the list again).

The information acquisition, decision-making, and action implementation functions of the task are set at the same level of automation for the DS condition as under the AS condition.

Communication between the AC and the IO in the task may include the same information as in the AS condition, depending on how much the team chooses to rely on the automation for setting priorities and how often the team cross-checks automation decisions. Teams may choose not to follow the AC’s requests for information via the interface and simply follow the target processing order provided by the automation, consequently reducing communication. However, additional communication between the AC and the IO is expected based on the need for operators to interpret the automation. For example, the automation will prioritize friendly targets at a very low level. This means that if the IO is following the order presented by the automation (and possibly ignoring requests from the AC) some friendly targets may reach the home base without information regarding their classification ever being displayed to the AC. This may result in questions and concerns for the AC.

6.1.1.4. Blended Decision Making

As a fourth automation condition, the TDT was modified to present the IO with a prioritized listing of targets, as in the DS condition, and to display a recommended classification for the target (see Figure 7). The IO display screen was modified to add a fourth column to the target and sensor data information section that displays an automated decision regarding the target classification. The target classification is determined using the same algorithm used for sorting the list. The automation considers the data reported by all three sensors and the reliability of those sensors at the time the report is given. The automation selects the aircraft type with the highest reliability. If the reliability is equal across the three sensor reports, then the automation selects the target reported by Sensor A. This classification is as good as the human team can be, except when all three sensor reports are equal. In this case, the human team can consider whether two different enemies or friendlies are reported and can consider the risk involved with respect to scoring points with a particular decision.

Figure 7. Section of IO display modified to show target classification for BDM.

This condition presents the same advanced automation of information analysis as the DS condition and the same degree of automation of information acquisition and action implementation as the DS and AS conditions. In addition, because an actual target classification is displayed to the IO, this condition provides a higher degree of automation of the decision selection function as part of the task than any of the other three conditions. Although classification decision information is provided to the operator, the operator must make the ultimate decision and must perform an action to implement this decision. This distinction separates the BDM condition from more advanced forms of automation such as Automated Decision Making or Supervisory Control as defined in Endsley and Kaber (1999).

As in the DS condition, communication between the AC and the IO in this condition is likely to depend on how much the team chooses to use the automation and how much they cross-check automation decisions. The automatic classification of targets is likely to limit communication between the AC and the IO regarding AWACS sensor positions for sensor reliabilities if the team chooses to trust the automation. Communication regarding priorities or target processing order should be similar to the DS condition.

6.1.2. Summary of Task and Operator Communications

The TDT provides a platform for evaluating the effects of different forms and levels of automation on teamwork. The goal of the task is to protect a friendly base from enemy attack. The AC and IO must work together to complete this goal. It is the role of the AC to monitor the radar screen and provide sensor reliability, target proximity, and target collision information to the IO. It is the role of the IO to use this information along with the sensor reports to classify targets as enemy or friendly and to communicate this classification to the AC. Based on this classification, the AC is expected to destroy enemy targets. There are three main areas in which the team members must communicate to complete their task. First, the AC must provide the IO with the reliability of the three AWACS sensors so that the IO can make an informed decision regarding the target classification. Second, the AC must communicate target proximity information to the IO so that the IO can prioritize targets for current classification. Third, the IO must provide the final target classification information to the AC in order for the AC to select a missile and shoot down enemies.

The four automation conditions created present four unique levels of automation that slightly alter the task and the communication requirements for the two team members. A summary of the task and communication requirements for each automation condition is presented in Table 8. The AS condition provides the lowest level of automation across the functions defined by Parasuraman et al. (2000). In this condition the AC must report sensor reliabilities to the IO. This may be done periodically, so that the IO is always aware of the current position of the AWACS as he or she considers each new target. Or, sensor reliabilities may be reported along with a specific target number for classification purposes. In this case, the AC may or may not adjust the report for a specific target based on the position of the sensor at the time that the target entered the display. Finally, sensor reliabilities may be reported only on request from the IO (for example, when the sensors report three different aircraft types). The AC must also provide target proximity information to the IO. This is generally done in one of two ways. One method is for the AC to call out each target in proximity order and the IO simply classifies targets in the order specified by the AC. A second method is for the IO to classify targets in any order he or she wishes, and the AC only calls out specific target numbers when they are close to the center or involved in impending collisions. Finally, the IO must provide target classification information to the AC. This can be done through the user interface by pressing a button and sending a color code or by verbalizing the target classification (or both). The IO may be specific about a classification (e.g., "Mig-29" or "enemy fighter") or may choose to make a simple general classification (e.g., "enemy"). In the latter case, the AC may have to shoot more than one missile to eliminate the target. However, this does not result in any scoring penalty or significant time loss so that only the general classification is needed for the team to perform adequately. The IO may also choose to verbalize details regarding the sensor reports; thus, sharing the classification decision with the AC.

The SHC condition provides an increased level of information acquisition and information analysis automation over the AS condition. In this condition, the communication requirements with respect to AWACS sensor reliability and target proximity are somewhat reduced due to additional information provided on both displays. However, because of the high visual load involved in the work of the IO and because some information is still unique to the AC (i.e., impending collisions), verbal communication by the AC may still be helpful in terms of optimizing performance for the team. In this condition, all of the methods of communication described for the AS condition are also possible as team strategies. In addition, the shared information may lead to additional communication between the team members regarding classification decisions.

The DS condition provides an increased level of information analysis automation over the AS condition. It provides the IO with a prioritized listing of target order based on the likelihood that a target is an enemy and the proximity of the target to the center. The requirement for sensor reliability reporting is similar to the AS condition. The AC is still required to report sensor reliabilities to the IO and may choose to do so in any of the ways described for the AS condition. The IO may use the knowledge that enemies are more likely to be sorted toward the top of his or her list to assist in decision-making and this may have some influence on the requests for and reporting of sensor reliabilities. The prioritized target order is different from the proximity order that is presented to the AC on the radar display. In addition, this prioritized listing is most useful when the IO is able to keep pace with the task and addresses all targets on the screen. In the case where some targets have not been classified and are about to reach the home base or collide, communication regarding target proximity is required from the AC. The AC may still choose to provide target proximity information in the same ways as described for the AS condition. Because the priority listing provided by the automation allows team members to focus on potential enemies first, performance of the team will be better if the IO follows the automated ordering with the AC calling out specific target numbers only when needed. With respect to communication of classification information, the same information is required as in the AS condition. Because the priority ordering of the list causes re-sorting that makes the interface slightly more difficult to use, this condition may induce more verbal reporting of target classifications by the IO.

Table 8. Communication requirements and team member roles by automation condition.

Condition	Communication Requirements and Team Member Roles		
	Sensor Reliability	**Target Proximity**	**Target Classification**
Action Support	AC must communicate sensor reliability to the IO verbally.	AC must communicate (either verbally or through the interface) to IO targets that are close to the center or involved in potential collisions. AC may choose to specify each target in proximity order.	IO must provide target classification to the AC either verbally or through the interface. The IO decision will be based on the sensor output combined with the sensor reliabilities given by the AC.
Shared Control	IO display includes report of current sensor reliability. IO task is visually loaded and reports of sensor position changes may be helpful from AC. AC may be more aware of sensor positions for specific targets at the time they entered the display.	IO display has a listing of the proximity of the targets to the center and may use this as a priority order for processing targets. AC should request specific targets that are close to the center or involved in impending collisions. AC may choose to specify each target in proximity order.	Same as AS except that when making the decision, the IO may view the sensor reliabilities on his or her own display screen.
Decision Support	AC must communicate sensor reliability to the IO verbally.	IO display presents targets listed in a priority order for processing. The order is different from a straight proximity order (which is supported by the AC radar display). The IO should follow the automation order with the AC requesting specific targets only when targets are close to the center or involved in an impending collision. AC may choose to specify each target in proximity order.	Same as AS except that the priority listing of targets may provide some additional information regarding classification since the IO knows that enemies are more likely to be near the top. In addition, the re-sorting of the list may induce more verbal reporting of classifications than the AS and SHC conditions.
Blended Decision-Making	IO display includes a classification by the automation that considers sensor reliabilities. AC should report sensor reliabilities, particularly in the case when all three sensors are equal and the automation may be unreliable.	Same as Decision Support condition.	IO must provide target classification to the AC either verbally or through the interface. The IO decision may be based on the automation, on the sensor data and sensor reliabilities given by the AC, or on some combination of the two. The re-sorting of the list may induce more verbal reporting of classifications.

The BDM condition provides an increased level of decision selection automation over the DS condition. This condition provides an automated classification of the target type that may result in reduced communication regarding sensor reliabilities, although the AC may provide sensor reliability information in any of the ways described for the AS condition. The IO may base the classification decision on the sensor reports and reliabilities, the classification reported by the automation, or some combination of both. Because the automation provides an essentially random report when the sensor reliabilities are equal, some reporting of sensor reliabilities is still required for the team to perform well. With respect to target proximity reporting, the BDM condition is the same as the DS condition. The best strategy is for the IO to follow the automated order and for the AC to report specific targets only when they are close to the center of the display or involved in an impending collision. Re-sorting of the list of targets may induce more verbal reporting of target classifications by the IO.

6.2. Experimental Design

A 4 x 2 mixed design was used in this experiment, with the 4 levels of automation as a between-subjects variable and 2 levels of task difficulty as a within-subjects variable. Ten teams were assigned to each automation condition. Subjects were randomly assigned to automation conditions and team member roles. Level of difficulty was determined by the number of targets present in the environment (on the task displays) at any given time: 5 targets for a low difficulty condition and 8 targets for a high difficulty condition. Bolstad and Endsley (1999b) used 3, 6, and 9 targets to represent low, medium, and high levels of difficulty in a study using the TDT. With these settings, they found significant performance and communication differences due to difficulty, particularly between the lowest and the two higher levels of difficulty. In addition, Young (1969) has shown that expert air traffic controllers can handle at most 9 to 10 targets at any given point in time (in terms of general information processing functions, the TDT is similar to air traffic control). Pilot testing of this version of the TDT revealed that a task condition of 9 targets was extremely difficult, making it almost impossible for teams to keep pace with the task. Since differences in automation conditions and crew coordination measures were expected to be greater at higher levels of workload (Clothier, 1991; Costley, Johnson, and Lawson, 1990), a high difficulty condition of 8 targets was used in this study. Since the difference between target levels of 6 and 9 was not strong in the Bolstad and Endsley (1999b) study, a condition of 5 targets was used to represent the lower level of difficulty in the present experiment.

Teams were presented with two trials in each experimental condition. In order to balance the order of presentation of level of difficulty, they experienced the first two trials as either low difficulty then high difficulty or as high difficulty then low difficulty. The second two trials were presented in the opposite order. The order of the presentation of task difficulty was balanced across teams (see Table 9). Order was considered as an additional nested variable in the experimental design. If analyses considering order as an independent variable indicated that the counter-balancing was effective (order was insignificant), then the order was dropped from further analyses.

Table 9. Experimental conditions.

	Automation Level							
	Action Support		Shared Control		Decision Support		Blended Decision-Making	
	Order							
	1	2	1	2	1	2	1	2
	Team Number							
	1, 9, 17, 25, 33	5, 13, 21, 29, 37	2, 10, 18, 26, 34	6, 14, 22, 30, 38	3, 11, 19, 27, 35	7, 15, 23, 31, 39	4, 12, 20, 28, 36	8, 16, 24, 32, 40
Trial	**Task Difficulty**							
1	Low	High	Low	High	Low	High	Low	High
2	High	Low	High	Low	High	Low	High	Low
3	High	Low	High	Low	High	Low	High	Low
4	Low	High	Low	High	Low	High	Low	High

Note: Low = 5 targets; High = 8 targets.

6.3. Subjects

Ten teams of two persons were assigned to each automation condition. Thus, 40 teams, or 80 subjects, were involved in the experiment. Subjects included undergraduate and graduate students at North Carolina State University, who participated voluntarily for monetary compensation ($7.50/hour). All subjects were required to have 20/20, or corrected to normal vision and some personal computer (PC) experience with a direct manipulation/graphical user interface. Subjects were required to speak and understand English fluently. Male-female, male-male, and female-female teams were partially balanced such that a similar number of each team makeup was presented in each automation condition. Teams that consisted of individuals who did or did not know each other were also balanced such that an equal number of each team type was represented in each automation condition. The effects of gender or inter-personal relationships were not an objective of this experiment. These precautions were taken simply to balance the composition of teams across automation conditions.

6.4. Apparatus

Team members sat at adjacent workstations separated by a partition so that they could not see each other (see Figure 8). Contemporary PCs were used to present the TDT. The computers were connected via an Ethernet-based Local Area Network. Each team member viewed his or her display on a 21-inch monitor with a 1024 x 768 pixel display resolution. A standard PC keyboard and a 2-D mouse were integrated with each computer to allow teammates to interface with the TDT. Team members communicated with each other using a FM radio headset, walkie-talkie. The headset covered one ear and a press-to-talk button was used for transmitting. All TDT operations could be completed with one hand so that the other hand was free for pressing the radio talk button in order to transmit. Although the subjects could hear each other in the test room if they spoke loudly, the radio output in the ear was compelling such that the radios were used consistently. A video camera was set up behind the two test participants to record the ongoing task activity on the two computer displays and to record test participant communications.

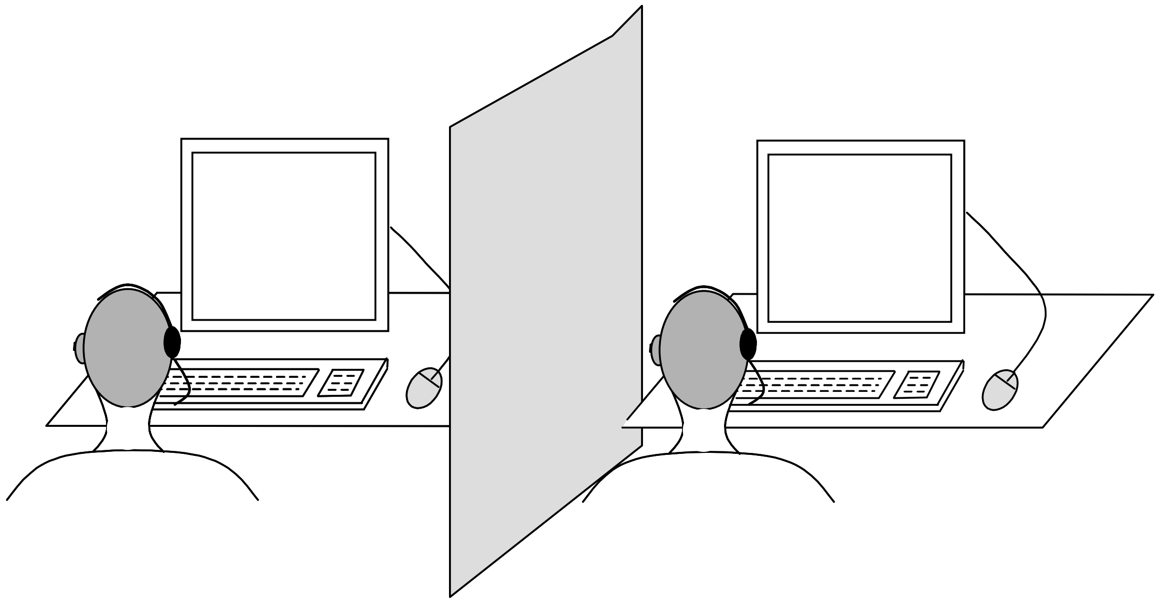

Figure 8. Experimental setup for test participants.

6.5. Dependent Measures

The review of literature on the measurement of teams has clearly indicated a need for assessment of both team effectiveness or outcome measures and team process measures. For this experiment, several measures were used to quantify the affect of automation on teams.

6.5.1. Team Effectiveness Measures

Team effectiveness was measured using the TDT point scoring system previously discussed. Points are scored when either a target is destroyed or it reaches the home base. Target destruction points are labeled *target points* and are positive for the elimination of enemy aircraft and negative for the destruction of friendly aircraft. The points are scaled such that a greater reward is given for enemy aircraft that may cause more destruction (e.g., fighters) and a greater penalty is associated with destroying friendly aircraft that are likely to be carrying more people (e.g., transports). A high score on *target points* represents good team performance.

Points assessed when an enemy target reaches the home base are known as *damage points*. Damage points are greater for aircraft that are likely to cause greater destruction. A low score on *damage points* represents good team performance. A combined team effectiveness measure can be determined by subtracting the *damage points* from the *target points*. Table 5 presents the actual values associated with target and damage points for the various types of aircraft presented in the TDT.

The number of target and damage points scored is dependent on the number of targets presented. The high difficulty condition of 9 targets provides more scoring opportunity than the low difficulty condition of 5 targets. Therefore, in order to remove the effect of the number of targets on the score, team effectiveness was also measured as a ratio of *target points, damage points,* and *total score* to the number of targets presented. The *target points ratio* provided a measure of the team's decision-making effectiveness for a single target. The *damage points ratio* provided a measure of the ability to keep pace with the task and, to a lesser extent, the team's decision-making. With respect to the team's decision-making, the damage ratio provides insight into whether subjects decided to let an enemy target pass through missile defenses.

The ability to score points is highly dependent on the reliability of the three AWACS sensors. Since the AWACS positions and movements are random, it is possible that for any given trial, the reliability of the sensors may be better or worse than in another trial. In order to assess the reliability of the AWACS sensors, an average reliability for each of the three sensors during the course of a trial was computed. The data for each trial was recorded on a per target basis, including the sensor outputs, sensor reliability, target types, target points, and damage points. The average reliability for the three sensors over a trial was calculated by summing the percent reliability for Sensors A, B, and C for each target throughout a trial and then dividing this number by the total number of targets multiplied by three (sensors). Pilot testing indicated that the average reliability of the sensors over a trial was 48% and varied as much as 10% from one trial to the next. For this reason, the average reliability of the sensors for a trial was calculated in order for it to be considered as a potential covariate in the analysis of the team effectiveness measures.

6.5.2. Team Coordination and Communication Measures

There is a great deal of evidence to support the importance of communication between team members for effective team performance (Billings and Reynard, 1984; Helmreich and Foushee, 1993). Two methods of measuring team communication and coordination have been discussed. One method is to categorize communication types and then quantify communication in accordance with the categorization (Costley, Johnson, and Lawson, 1989; Urban et al., 1993). A second measure is to evaluate the quality of communication by rating teams across categories that have been shown to be important to the overall effectiveness of teams (Bowers et al., 1992; Brannick et al., 1995). This study used both methods to assess team coordination.

Pilot testing was conducted to identify the types of communications that team members used. Transcripts of videotape recordings were analyzed for specific communication types that could be quantified. Based on these transcripts, the following communications were identified:

- Target classification requests from the AC, with or without a response from the IO
- Unsolicited target classification reports by the IO
- Sensor reliability requests from the IO, with or without a response from the AC
- Unsolicited sensor reliability updates from the AC
- Strategy suggestions initiated by the IO or AC
- Clarifications initiated by the IO or AC
- Other communications initiated by the IO or AC

Target classification requests refer to the calling-out of specific target numbers by the AC for classification by the IO. A classification request without a response does not necessarily mean that the IO ignored the request from the AC, just that they made no verbal response. In most cases this indicates that the IO chose to respond to the AC using the user interface (by color coding the targets) rather than respond verbally.

During pilot testing, it was observed that much of the communication between members regarding strategy and decision-making occurred prior to, or between, trials.

Therefore, the experimental procedure was modified to give teams a 5-minute period prior to each trial for strategy discussion (and to prohibit these communications at other times), which could then be included in the team coordination data collection. Therefore, the following communication was added to the list of those to be counted:

- Pre-trial strategy comments initiated by the IO or the AC

Two raters counted the communications during a trial. Counts were recorded on the data sheet shown in Appendix A. Raters simply marked a tally for each type of utterance. A single mark was recorded for an entire passage that represented each of the categories. For example, if the AC said "Give me 158, Sensor C is best" and the IO responded "158 is friendly", one tally would be recorded for *target classification request from the AC with IO response* and one tally would be recorded for *unsolicited sensor update from AC*. One rater made the counts in real-time during the trial and the second rater used videotape recordings of the trial. The counts based on the videotape were also completed in real-time (without rewinding and listening again) so that the experience and level of accuracy was the same for both raters. Correlations on pilot count data indicated high inter-rater reliability with r greater than .90 for the four most common types of communication.

Beyond communication counts, an additional measure of team communication was determined; an *anticipation ratio* (Serfaty, Entin, and Johnston, 1998) was calculated by summing the tallies that represented some type of information transfer between team members, summing the tallies that represented information requests, and then calculating the ratio of the information transferred to the information requested.

Based on the inconsistent evidence of quantity of communication as an indicator of good teamwork (Orasanu, 1990; Wiener, 1993), this study also used outside observer ratings of team coordination quality as a measure of team processes. The measurement of team coordination follows the conceptual framework of Dickinson and McIntyre (1997) that proposes development of observation (or event) scales that indicate specific behaviors associated with high or low performance along several dimensions of teamwork, followed by recording team performance on these behaviors to provide a basis for a final team rating. This rating is made on a five-point scale for each of the teamwork dimensions. The teamwork dimensions used in this study are similar to those used by Brannick et al. (1995) and those defined for the ACOE (Bowers et al., 1992). They include:

(1) assertiveness,
(2) decision-making,
(3) situation assessment,
(4) leadership, and
(5) communication.

In using this method, two raters observed team members during performance of the experimental trials, or from videotape recording of the trials. The rater using the videotape watched the team practice sessions so that he was aware of any teamwork behaviors, such as the development of standard communications, which may have occurred during practice. The raters recorded observations of good or poor teamwork on each of the five teamwork dimensions on the same sheet that was used for counting communications (see Appendix A). Raters studied definitions of each of the five dimensions of teamwork and example behaviors considered to represent skill in each of these dimensions (see Table 10). Based on these definitions, they marked a tally in the column for each teamwork dimension when they observed a behavior that was associated with good or poor skill in that dimension. Comments were also recorded as a memory aid for end-of-trial ratings and as qualitative information for interpreting the results of the experiment.

At the end of a trial, the raters reviewed the tally marks for each of the teamwork dimensions and recorded an overall rating on each of the six team-coordination dimensions. Raters considered both the number of times teamwork behaviors were exhibited, as well as the quality of the behaviors, in their final rating. Scales for each of the teamwork dimensions were provided to raters as a reference (see Figure 9 for an example scale on "decision-making"). (All of the scales used in the team ratings are included in Appendix A.) The team coordination ratings resulted in one overall rating per team on a scale from 1 ("hardly any skill") to 5 ("complete skill") along each of the six dimensions of teamwork for each trial.

Table 10. Teamwork definitions and associated behaviors.

Adapted from: Bell and Lyon, 2000; Bowers, Braun, and Morgan, 1997; Bowers, Morgan, Salas, and Prince, 1993; Canon-Bowers et al., 1995; Dickinson and McIntyre, 1997; Kaber and Endsley, 1997.

<table>
<tr>
<td>

ASSERTIVENESS
Assertiveness refers to the willingness to make decisions, demonstrating initiative, and maintaining one's position until convinced otherwise by the facts.

Behaviors that suggest assertiveness include:
• Confronting ambiguities and conflicts
• Asking questions when uncertain
• Maintaining a position when challenged
• Making suggestions
• Stating an opinion on decisions, procedures, or strategies

DECISION-MAKING
Includes identifying possible solutions to problems, evaluating the consequences of each alternative, selecting the best alternative, and gathering information needed prior to arriving at a decision.

Behaviors that suggest decision-making skill include:
• Communicates possible solutions
• Gathers information to evaluate solutions
• Communicates consequences of alternatives
• Cross-checks information sources
• Selects the best alternative
• Development of plans
Implements the decisions that were made

LEADERSHIP
Team leadership involves providing direction, structure, and support for other team members. It does not necessarily refer to a single individual with formal authority over others. Team leadership can be shown by several team members.

Behaviors that suggest leadership skill include:
• Explains to the other team member exactly what is need from them during the task
• Listens to the concerns of the other team member
• Provides statements of team direction, strategy, or priorities for the task
• Sets goals for the team and orients the team toward those goals
• Provides feedback to the other team member regarding his/her performance

</td>
<td>

COMMUNICATION
Involves the exchange of information between two or more team members in the prescribed manner and by using proper terminology. One purpose of communication is to clarify or acknowledge the receipt of information.

Behaviors that suggest communication skill include:
• Verifies information prior to taking an action
• Acknowledges and repeats messages to ensure understanding
• Uses accurate terminology
• Makes concise statements with little extraneous information
• Establishes and uses conventional or standard speech (e.g., acronyms/shortcuts)
Provides unsolicited responses (gives more detail than was asked, when appropriate)

SITUATION ASSESSMENT
Situation assessment refers to the verbalization of information related to the perception of elements in the environment, the comprehension of their meaning in terms of task goals, and the projection of their status in the near future. Situation assessment verbalizations may serve to promote shared situation awareness between team members.

Behaviors that suggest situation assessment include:
• Situation assessment updates in which team members communicate the current state of the system
• Identification of problem situations and recognizing the need for action
• Exchange of information for the prevention of errors
• Noting deviations in SA between team members
• Demonstrated awareness (e.g., via verbal communication) of the ongoing mission status and the overall goal
• Integration of information from multiple sources
• Accurately prioritizing information and actions

</td>
</tr>
</table>

DECISION-MAKING Includes identifying possible solutions to problems, evaluating the consequences of each alternative, selecting the best alternative, and gathering information needed prior to arriving at a decision. Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of decision-making:		
Complete skill in decision-making	5	Team members gather information to evaluate possible solutions. Team members always evaluate each solution and explore consequences of the solutions. Team members cross-check information sources. Based on the information and consequences, team members select the best alternative. Team members periodically plan their activities and always follow the plans and decisions made.
Very much skill in decision-making	4	
Adequate skill in decision-making	3	Team members sometimes gather information to evaluate possible solutions. Team members sometimes explore the consequences of the solution. Team members rarely cross-check information sources. Based on the information and consequences considered, team members usually select the best alternative. Team members sometimes plan activities and usually follow the plans and decisions made.
Some skill in decision-making	2	
Hardly any skill in decision-making	1	Team members rarely gather information and explore the consequences of the solution. Decisions are made arbitrarily with little consideration of the information available. Team members rarely plan activities and may not follow the plans and decisions that are made.

Figure 9. Team coordination dimension rating scale.

Raters practiced conducting ratings during pilot testing. Following these practice sessions, the two raters met and discussed specific behaviors associated with the TDT that they believed were representative of specific teamwork dimensions and agreed on specific behaviors and guidelines. For example, with respect to situation assessment, frequent updates of sensor reliabilities and verbalization of target numbers that were nearing the home base was considered the average or standard level of situation assessment skill and resulted in a rating of 3. In order to receive ratings of 4 or 5, teams had to verbalize team performance measures (e.g., score), targets associated with impending collisions, and/or other situation assessment observations such as “all targets are friendly now”. The frequency and variety of these observations determined specific ratings of 4 and 5.

Brannick et al. (1995) evaluated a similar rating method using multitrait-multimethod analysis and found good convergent and discriminant validity between judges. Correlations between raters in other studies using this method have ranged from $r = 0.55$ to 0.87 (Brannick et al. 1993); $r = 0.48$ to 0.71 (Travilian et al., 1993); and $r = 0.55$ to 0.97 (Volpe et al., 1996). These studies suggest that ratings with inter-rater reliabilities of 0.45 and greater are acceptable for further analysis. Pilot testing as part of the present research resulted in slightly lower inter-rater reliabilities ranging from $r = .35$ for decision-making to $r = .85$ for situation assessment. Based on these results, further training of raters was completed prior to conducting experimental trials. Because Brannick et al. (1995) found poor convergent validity between the same teams under different scenarios, this study involved two trials per team in each experimental condition and separate ratings were scored for each trial.

6.5.3. Subjective Workload Measure

In addition to measuring team effectiveness and team processes, both team members rated perceived workload. Team members were asked to complete two separate workload ratings, one to assess *taskwork* and another to assess *teamwork*. Definitions of the concepts of *taskwork* and *teamwork* are provided in Table 11.

Workload was measured using a modified NASA-TLX scale (Hart and Staveland, 1988). In using the NASA-TLX measure, subjects rank and rate six dimensions of workload including: (1) mental demand, (2) physical demand, (3) temporal demand, (4) performance, (5) frustration, and (6) effort. The rating of each individual dimension is then multiplied by a weighting factor and the results are summed to create an overall workload score. The weighting factors are determined by having subjects make paired comparisons of the various workload dimensions in terms of which is more important for the task at hand. Team members completed two rankings, one for taskwork and one for teamwork, following a practice trial. These weighting factors were used for all subsequent ratings. Team members rated both taskwork and teamwork following each trial, resulting in 16 weighted workload scores (4 trials x 2 workload types x 2 team members) for each team. Details of the procedure for collecting the workload data are presented in the procedures section.

Table 11. Definitions of taskwork and teamwork for rating purposes.

Taskwork
Taskwork refers to the workload associated with the individual tasks each team member must perform. Examples of taskwork within the Theatre Defense Task include: (1) monitoring the displays for information, (2) using the mouse and keyboard to work with the system, (3) analyzing the information displayed to make decisions required specifically for your role in the mission, etc.
Teamwork
Teamwork refers to the interactions between team members that are necessary for exchanging information, developing and maintaining communication patterns, coordinating actions, and negotiating decisions. Examples of teamwork within the Theatre Defense Task include: (1) communicating verbally with your team member, (2) communicating via the user interface with your team member, (3) coordinating the timing of actions with your team member, (4) working with your team member to develop communication standards or task strategies, etc.

6.5.4. Background Questionnaire

In addition to the measures described above, a short questionnaire was given to each team member to assure that subjects met the qualifications of the study and to record background information. The survey is shown in Appendix B and included subject sex, age, video-game experience, and PC experience.

6.6. Procedure

The procedures for the experiment included the following steps:

(1) Introduction to the experiment, including completion of consent forms, background questionnaire, and payment form (if applicable);

(2) Training on the TDT;

(3) One 15-minute practice session;

(4) Completion of a practice rating of taskwork and teamwork using the NASA-TLX scale and completion of the NASA-TLX paired comparisons for taskwork and teamwork;

(5) Four 15-minute Experimental trials; and

(6) Each trial was followed by taskwork and teamwork workload ratings using the NASA-TLX rating scale after each trial.

Short breaks were given as needed following the training session, the practice trial, and each of the experimental trials. Longer (10-minute) breaks were given following the practice trial and between Trials 2 and 3. The entire experimental session lasted between 2.5 and 3 hours. All of the forms that were used in the experiment are shown in Appendices A and B. Appendix C provides a full transcript of the instructions that were presented verbally to the team members throughout the experiment, including training in the task.

Training was given to the team members for their specific automation condition. Team members were informed that, although they would ultimately be performing the task in different locations, they had come together in the same location for training so that they would be fully aware of the tasks of the other team member. The experimenter first described the equipment for the tasks, and the information presented on the AC and IO displays. The experimenter then instructed the subjects in the task from the point in time at which an aircraft appeared on the screen, to the classification of the aircraft, to the elimination of enemy aircraft. All of the details regarding the automation, point-scoring (including the unique scoring of collisions), and sensor readings were provided during this session. Particular emphasis was given to the fact that sensor reliability was dependent on the distance to the center and not to the proximity to the aircraft (a common misconception) and that the sensor readings were provided at the time the aircraft entered the screen and were not updated based on sensor movement during the time that the aircraft traveled toward the home base.

After presenting all of the details of the task, the experimenter described to the team members the types of information they might choose to communicate verbally, including aircraft proximity to the home base, impending collisions, sensor reliability, and sensor classifications. The experimenter also provided the team members with a list of standard terms that may be used for communication (included in Appendix C). Subjects were informed that they could use these terms or, if they were uncomfortable using them, they could develop their own terms. Finally, the experimenter presented three hints that helped the team to understand the urgency of the task, the value of specific information with respect to point scoring, and the difference in workload between the IO and the AC. For example, with regard to point scoring, only the enemy/friendly classification is needed and exact correctness does not improve the overall score.

After the verbal presentation of training by the experimenter, the team members were given 5 minutes to “try out” the task in their role, using the radios to communicate. After 5 minutes, they were asked to trade seats with the other team member and were given 5 minutes to perform the task in their team member’s role. This additional cross-training experiment gave the team members a better understanding of both roles so that they could more fully understand the communication requirements (Travillian et al., 1993; Volpe et al., 1996).

After the training was completed, subjects were shown the definitions of the NASA-TLX workload dimensions and definitions of taskwork and teamwork as presented in Table 11. They were told that they would be asked to make ratings along the six workload dimensions for both taskwork and teamwork following the practice session and following each experimental trial. They were also informed of the observer ratings of teamwork skills and were asked to refrain from speaking about the task except during a 5-minute period prior to each trial (including the practice trial) and during the trials. The entire introduction and training session took a little over an hour.

Team members were then asked to imagine that they were in different locations and to only use the radios for communication. They were given 5 minutes to discuss their team strategy for performing the task prior to the start of the 15-minute practice session. When the team members were finished with their discussion, they were asked to press the appropriate buttons to begin the task. For the practice session, the task was run in the low difficulty level condition. A minimum performance criterion for the practice session was established to identify and ensure adequate skill and training for each of the teams. The criterion considered the total number of enemy kills for the practice trial and required teams to perform within 2 standard deviations of the mean performance of subjects completing the practice in pilot testing.

Following the practice trial, subjects rated first the workload associated with their taskwork, and then the workload associated with their teamwork. This practice rating was intended to allow the subjects to become familiar with the scale and with rating both taskwork and teamwork. Throughout all of the ratings the definitions of the workload dimensions and the definitions of taskwork and teamwork were available to the subjects. All ratings were conducted with taskwork being evaluated first and teamwork second. After completing the practice ratings, the subjects completed the NASA-TLX paired comparisons.

After the practice session, subjects completed the experimental trials. Each team completed four 15-minute trials, two of each of the two levels of task difficulty, as shown in Table 9. Each trial was preceded by the 5-minute strategy discussion period and followed by the NASA-TLX rating of teamwork and taskwork load.

7. HYPOTHESES

7.1. Level of Automation Effects

It was hypothesized that level of automation would affect team effectiveness scores, team communication counts, team coordination ratings, and workload ratings. With respect to team effectiveness scores, performance was expected to be greater with increasing automation. That is, target points should be higher and damage points should be lower for SHC, DS, and BDM compared to the AS condition. In addition, target points should be higher and damage points should be lower for the BDM condition compared to the DS and SHC conditions. This is based on the assumption that the automation of information acquisition, information analysis, and decision selection provides added information and performance enhancing features for the task. Automation is implemented in the TDT such that it provides high quality information to operators. The algorithms used to prioritize targets for processing and classifying represent good decision aiding with respect to the goals of the task. In the comparison of the SHC and the AS conditions, this hypothesis is supported by the results of Bolstad and Endsley (1999b, 2000) and the results of this experiment were expected to replicate their findings.

With respect to team communication counts, lower counts were expected for the SHC condition compared to all other conditions. This is because the SHC condition provides team members with necessary information directly on their displays. In particular, counts related to the AC providing reliability information to the IO were expected to be lower for the SHC condition.

With respect to team coordination ratings, higher ratings were expected with automation of the information acquisition and information analysis aspects of the task, as compared to automation of decision selection. There are several reasons for this prediction. First, Kaber et al. (1999) have shown that intermediate forms of automation result in better situation awareness for individuals than more advanced forms of automation. Improved situation awareness for individuals should also lead to improved situation awareness for the team. Second, automation of information acquisition and information analysis was expected to support team members in attaining situation awareness through the presentation of task specific information on the displays. This, in turn, was expected to enhance shared situation awareness (or lead to common mental models) since the displays presented related information to both team members. A third reason for higher ratings of team coordination associated with intermediate levels of automation (automation of information acquisition and information analysis) is that the team members are required to consider the information provided by the automation to make decisions. This activity was expected to keep them engaged in the task and to require them to coordinate and share information that would also support better shared situation awareness. Automation of the decision selection aspect of the task, however, was expected to result in team members blindly following the recommendations of the automation, perhaps without any understanding of the rationale behind the decision, or with different understanding developing between team members, possibly resulting in complacency. Conversely, automation of the decision selection aspect of the task may simply be ignored and, thus, serve only as a distracter on the display.

Of the team process behaviors measured, situation assessment, decision-making, leadership and communications were expected to be influenced by the various automation conditions. Specifically, ratings of communications, leadership, and decision-making were expected to be higher in the SHC condition than in all other automation conditions. This is because the SHC condition provides moderate automation of both information acquisition and information analysis. This was expected to lead to common knowledge between team members which should support more efficient and standard communications, better discussion regarding decisions, and more supportive behaviors that would be reflected in leadership ratings. Ratings of situation assessment however, were expected to be lower for the SHC condition than for all other conditions, since there was less need for team members to communicate situation assessment information. Ratings of situation assessment, communications, leadership, and decision-making were expected to be higher in the AS condition than in the BDM condition. That is, coordination was expected to be better with very little automation than with the most advanced form of automation since the BDM condition was expected to introduce complexity that would negatively affect team members' coordination. Comparisons between the DS condition and the AS and BDM conditions were difficult to predict. While the DS condition represented automation of information analysis, it provided a relatively high degree of automation (prioritized listing of targets) and did so in the absence of advanced presentation of information. This condition was expected to result in team coordination ratings similar to those of the BDM condition and worse than the AS condition. It was expected that team members might not understand the logic underlying the automation and might become complacent and simply follow the decisions recommended based on the information analysis.

With respect to the workload ratings, taskwork and teamwork ratings were expected to generally decrease with increasing automation. With increasing automation, team members are required to do less of the taskwork and teamwork on their own and can rely on the automation to perform tasks and present information they may otherwise have to obtain from a team member. However, some researchers suggest that decision-aiding conditions may increase mental workload due to a need to evaluate the system's advice (Harris et al., 1993; Kaber and Riley, 1999; Selcon, 1990). Therefore, the effects of the automation conditions on workload depend on operator trust and understanding of the automation. For this study, taskwork and teamwork load ratings were hypothesized to be lower for the BDM condition than for all other conditions, under the assumption that team members would trust the automation and would not spend undue time evaluating the decisions of the automation. Taskwork and teamwork load ratings were hypothesized to be lower for the DS condition than for the SHC and AS conditions. However, the SHC condition was not expected to produce lower workload than the AS condition. Since the SHC condition provided additional displays to the user with information they would otherwise obtain from a team member, this condition was expected to result in a decrease in teamwork but an increase in taskwork as compared to the AS condition.

7.2. Task Difficulty Effects

Task difficulty was expected to affect team effectiveness scores, workload ratings, and team coordination ratings. With respect to team effectiveness, target points were expected to be higher in the high difficulty condition than in the low difficulty condition. This is because more targets were presented to the teams in the high difficulty condition, providing a greater opportunity for scoring target points. However, target ratio, which reflects the accuracy of the decisions for a single target was expected to be higher for the low difficulty condition, since team members would have more time to evaluate the information presented. Damage points were also expected to be higher in the high difficulty condition, since more enemy targets were presented, more could reach the home base. Damage ratio was expected to be lower for the low difficulty condition, reflecting a better ability of teams to keep pace with the task.

With respect to workload ratings, the high difficulty condition was expected to result in higher ratings of both taskwork and teamwork.

With respect to team coordination ratings, higher ratings were expected in the low difficulty condition than in the high difficulty condition. In addition, an interaction effect was expected due to the task difficulty and automation manipulations. All of the effects of team coordination due to automation conditions described in the previous section were expected to be apparent to a greater degree in the high difficulty condition than in the low difficulty condition. Previous research on teams and automation has shown that significant differences between designs tend to appear when workload is higher (Bolstad and Endsley, 2000; Clothier, 1991; Costley, Johnson, and Lawson, 1989). Under extreme workload conditions, it is believed that teams are required to rely more on teamwork skills; therefore, the effects of systems that support or inhibit team coordination are expected to be more apparent at high levels of task difficulty.

7.3. Summary of Hypotheses

Table 12 summarizes the hypothesized level of automation and task difficulty effects for all team effectiveness, team coordination and communication, and workload measures recorded during the experiment. This Table is subsequently referenced in the detailed discussion of the experimental results.

Table 12. Summary of experiment hypotheses.

Level of Automation Effects
Team Effectiveness Measures
H1: Lower target points and target ratio and higher damage points and damage ratio for AS compared to all other conditions
H2: Higher target points and target ratio and lower damage points and damage ratio for BDM compared to SHC, and DS conditions
Communication Counts
H3: Lower communication counts for SHC compared to all other conditions
H4: Lower communication counts related to reliability for SHC compared to all other conditions
Team Coordination Ratings
H5: Higher ratings of communication, leadership, decision-making, and total teamwork for SHC compared to all other conditions.
H6: Lower ratings of situation assessment for SHC compared to all other conditions
H7: Lower ratings of communication, leadership, decision-making, situation assessment and total teamwork for BDM compared to all other conditions.
H8: Lower ratings of communication, leadership, decision-making, situation assessment and total teamwork for DS compared to AS.
Workload Ratings
H9: Lower ratings of teamwork and taskwork for BDM compared to all other conditions.
H10: Lower ratings of teamwork and taskwork for DS compared to AS and SHC.
H11: Lower ratings of teamwork for SHC compared to AS
H12: Lower ratings of taskwork for AS compared to SHC
Task Difficulty Effects
Team Effectiveness Measures
H13: Higher target points and higher damage points for high difficulty compared to low difficulty
H14: Higher target ratio for low difficulty compared to high difficulty
H15: Lower damage ratio for low difficulty compared to high difficulty
Workload Measures
H16: Higher ratings of teamwork and taskwork for high difficulty compared to low difficulty
Team Coordination Ratings
H17: Higher team coordination ratings in low difficulty compared to high difficulty.
Automation Condition and Task Difficulty Interactions
H18: Level of automation effects on team effectiveness and team coordination more pronounced in the high difficulty condition compared to the low difficulty condition.

8. DATA ANALYSES

For the analysis of team effectiveness and workload measures, a three-way analysis of variance (ANOVA) model was used with difficulty as a within-subjects variable and automation condition and order as between-subjects variables. The ANOVA also included appropriate subject and error terms, and the interactions between difficulty, order, and automation condition:

$$Y = LOA + ORDER + DIFF + TEAM(LOA*ORDER) + LOA*DIFF + LOA*ORDER + DIFF*ORDER + TEAM(LOA*ORDER)*DIFF + LOA*DIFF*ORDER + \varepsilon$$

In the cases where this model did not result in any significant effects due to order, analyses were conducted with a reduced model:

$$Y = LOA + DIFF + TEAM(LOA) + LOA*DIFF + TEAM(LOA)*DIFF + \varepsilon$$

Order effects were expected to be insignificant due to counterbalancing. Any significant effects revealed by the ANOVAs were further analyzed using Duncan's Multiple Range test with an alpha level of 0.05. The following sections detail the data handling and the specific analyses conducted on each of the dependent measures. Unless otherwise specified, reference to an ANOVA refers to the models described above.

Residual analyses were conducted to ensure that the underlying assumptions of normality and constant variance of the ANOVA were upheld by the data sets. Residual plots, normal probability plots, and normality statistics (the Shapiro-Wilks test) were used to verify these conditions. The outcomes of these analyses for specific response measures are discussed in the following subsections.

8.1. Team Effectiveness Measures

The six team effectiveness measures mentioned previously include:

- Target score,
- Target score ratio,
- Damage score,

- Damage score ratio,
- Total score,
- Total score ratio.

Target and damage scores as well as the total number of targets presented to subjects were recorded by the TDT program in a data file at the end of each trial. Total score was calculated by subtracting the damage score from the target score. Each of the ratio measures were calculated by dividing the raw score by the total number of targets. The six team effectiveness dependent measures resulted in six separate data sets of 160 observations each (40 teams x 2 levels of difficulty x 2 trials).

The reliability of the AWACS sensors affects a team's ability to accurately classify a target. Since there was some variability in average sensor reliability during a trial, correlations analyses were conducted on the average sensor reliability and each of the team effectiveness measures to determine whether this variability affected the final team performance. Based on these correlations, all of the team effectiveness measures correlated significantly ($p<0.05$) with average sensor reliability, as shown in Table 13. Therefore, all six team effectiveness measures were analyzed using the ANOVA models described above with the addition of a covariate, *reliability,* as a random effect.

The order effect did not prove to be significant for the analyses of target score, target score ratio, total score, and total score ratio and, consequently, the reduced model was used (including the reliability covariate) for the final analysis of these measures. With respect to the damage score and the associated ratio, the trial order was significant; therefore, the full ANOVA model (including the reliability covariate) was used for these analyses.

Table 13. Correlation of team effectiveness measures with average sensor reliability.

Team Effectiveness Measure	Pearson Correlation Coefficient, r (N=160)	Prob > H_0: $\rho = 0$, P
Target score	0.50	< 0.0001
Target score ratio	0.59	< 0.0001
Damage score	-0.27	< 0.001
Damage score ratio	-0.34	< 0.0001
Total score	0.55	< 0.0001
Total score ratio	0.59	< 0.0001

8.2. Team Coordination Measures

8.2.1. Team Communication Counts

Fourteen different measures (counts) of team communication were recorded during each test trial including:

1. Target classification request from AC with response from IO;
2. Target classification request from AC without response from IO;
3. Unsolicited target classification reports by IO;
4. Sensor reliability request from IO with response from AC;
5. Sensor reliability request from IO without response from AC;
6. Unsolicited sensor reliability update from AC;
7. Strategy suggestion initiated by IO;
8. Strategy suggestion initiated by AC;
9. Clarification initiated by IO;
10. Clarification initiated by AC;
11. Any other communication initiated by IO;
12. Any other communication initiated by AC;
13. Pre-trial strategy comment made by IO; and
14. Pre-trial strategy comment made by AC.

However, the number of observations of *sensor reliability requests without a response* was very low (300 out of 320 trials produced a count of 0) because the AC almost always responded to requests from the IO. Therefore, this measure did not appear to represent an important part of communication in the task and was dropped from the overall analysis. In addition, the number of strategy suggestions, other comments, and clarifications that occurred during a trial, as well as the number of pre-trial strategy comments were very low. Because strategy suggestions, other comments, and pre-trial strategy comments tended to all be related to discussion of the task and/or situation assessment statements, these measures were combined to create an *other communications* measure that generally reflected task-related situation assessment, decision-making, or strategy statements by either the AC or the

IO. Since clarifications did not appear to provide any meaningful information, the counts also were dropped from the overall analysis. Finally, an additional measure added to the overall analysis was the sum of all of the communications for a trial. The resulting set of communication counts that were analyzed included:

1. Target classification request from AC with response from IO;
2. Target classification request from AC without response from IO;
3. Unsolicited target classification reports by IO;
4. Sensor reliability requests from IO with response from AC;
5. Unsolicited sensor reliability update from AC;
6. Other communications; and
7. Total communications.

Thus, seven data sets of communication counts, each including 160 observations per rater or 320 total observations, were statistically analyzed. Correlations were conducted on pairs of data sets produced by the raters to verify inter-rater reliability on the counts.

Residual analyses of the normality and homogeneity of variance of the communication count data indicated that, for the majority of the measures, the communication data were non-normal. For some of the measures, there was a strong floor effect (many observations of zero) and transformations of the data could not be used to resolve the potential ANOVA assumption violations. Consequently, non-parametric analysis of the data based on ranks was elected as an alternate methodology because non-parametric tests do not place the same requirements on the data as parametric tests. Non-parametric tests are also comparable in terms of identifying statistically significant differences among conditions. Since the preponderance of non-parametric tests are one-way analyses, investigation of the independent variables, automation condition and level of difficulty, were conducted separately.

With respect to automation condition, the data was divided into two separate data sets for analysis, one representing the low difficulty condition and the other representing the high difficulty condition. Since automation condition was a between subjects variable, the non-parametric analysis required the comparison of four independent samples. The Kruskal-

Wallis test allows for the comparison of several independent samples without the assumption of normality of the data (Conover, 1980). However, this test does not accommodate designs involving repeated measures. With respect to the present experiment, multiple trials under a single condition or multiple observations by different raters on a single condition could not be accounted for in this type of analysis. Consequently, the communication data was averaged across the two trials under each difficulty condition and the communication data was averaged across the two raters. This produced two data sets of 40 observations each for each of the seven communication measures. Since trial order was balanced within each automation condition, no effect of order was expected in this analysis.

Each data set was subjected to the Kruskal-Wallis test for several independent samples. For those communication count measures where the result was significant, separate Kruskal-Wallis tests were conducted on each of the pairs of automation conditions (e.g., AS vs. SHC, AS vs. DS, etc.) to determine which of the conditions were significantly different from each other. This is a commonly accepted approach to comparing conditions using the non-parametric test.

Visual inspection of the communication count data revealed differences between automation conditions that were not detected as statistically significant. Because of the large variability in the count data and the data reduction necessary for conducting the non-parametric tests, power was calculated for the seven communication measures. Power $(1 - \beta)$ was calculated using the method for an F test as part of a one-way ANOVA (Neter et al., 1990). Conover (1980) states that the efficiency of non-parametric tests based on ranks is about 95% of a similar parametric analysis. The results of the power calculations are shown in Table 14. For several of the communication types, the power of the statistical test was low. For *classification requests with a response* the power of the test was so low that the analysis is not meaningful; therefore, results are not reported. For other measures for which the power of the test was low (on the order of 0.20 to 0.60), a more relaxed interpretation of the statistical results was used and trends in the data ($p < 0.15$) are presented and discussed.

Table 14. Power of statistical test for detecting differences between automation conditions for specific communication types

	Power (1 – β)	
Communication Type	Low Difficulty	High Difficulty
Classification request from AC with response from IO	0.09	0.11
Classification request from AC without response from IO	0.28	0.48
Unsolicited classification from the IO	0.31	0.18
Reliability request from the IO with a response from the AC	0.67	0.48
Unsolicited reliability report by the AC	0.78	0.88
Other communications	0.36	0.52
Total communications	0.40	0.61

With respect to the difficulty variable, the data was separated into four different sets, one for each automation condition. Task difficulty represented a within-subjects variable; therefore the data included pairs of observations on each team: one score under low difficulty and one under high difficulty. For a single sample of pairs of observations, Conover (1980) suggests the Wilcoxon Signed Ranks Test. For this analysis, the communication data was again averaged across the two trials and two raters. For each communication count measure, there were four data sets of 20 observations each. A two-tailed Wilcoxon Signed Ranks Test was conducted on each data set (without making any predictions as to whether communication counts were expected to be higher or lower based on level of difficulty).

8.2.2. Anticipation Ratio

The *anticipation ratio* was calculated as an additional measure of team communication. The communication count data sets that were averaged across trials and raters were used in determining the anticipation ratio. Categories that were summed to create the total *information transfer* measure included the following:

- Target classification requests from the AC, with a response from the IO,
- Unsolicited target classification reports by the IO,
- Sensor reliability requests from the IO, with a response from the AC,
- Unsolicited sensor reliability updates from the AC, and
- Other communications.

Categories that were summed to create the total *information request* measure included:

- Target classification requests from the AC, with a response from the IO,
- Target classification requests from the AC, without a response from the IO, and
- Sensor reliability requests from the IO, with a response from the AC.

The anticipation ratio was calculated by dividing the information transfer measure by the information request measure. One team had no information requests for one of the two levels of difficulty. Thus, the anticipation ratio resulted in division by zero. For this condition, the maximum anticipation ratio of all other conditions (a value of 2) was used in substitution. The anticipation ratio was analyzed using the Kruskal-Wallis test for effects due to automation condition and by the Wilcoxon Signed Ranks test for effects due to level of difficulty.

8.2.3. Correlations Between Communication Counts and Team Effectiveness Measures

To determine if any of the specific types of communications tallied were associated with either high or low performance, linear correlations were conducted on the six performance measures and each of the seven communication types and the anticipation ratio. In addition, in order to determine whether any specific types of communications tended to occur together and to determine whether any of the communication types that did correlate with performance were also correlated with each other, linear correlations were conducted on the seven communication types.

8.2.4. Team Coordination Ratings

The team coordination ratings resulted in 160 observations per rater for each of the five dimensions of crew coordination. In addition, a measure of overall team coordination was obtained by summing the scores on the five dimensions for the two raters. A similar method has been employed for assessing overall team coordination in other studies (Volpe et al., 1996). A linear correlation of the ratings by Rater 1 and Rater 2 was determined for all five coordination dimensions and the overall coordination measure in order to assess inter-rater reliability.

Because the rating data consisted of scores of “1”, “2”, “3”, “4”, or “5” for each dimension of team coordination, the resulting data was discrete and based on statistical diagnostics the normality assumption of the ANOVA was violated. Consequently, non-parametric analyses similar to those used for the coordination data were conducted on the teamwork ratings.

With respect to the automation condition, the data was separated into two data sets, one for low difficulty and one for high difficulty. The data was then averaged across the two trials and the two raters. The resulting 12 (6 teamwork measures x 2 difficulty conditions) data sets of 40 observations each were subjected to the Kruskal-Wallis test for several independent samples. For those data sets that revealed a significant effect, additional Kruskal-Wallis tests were conducted to compare each automation condition with every other automation condition.

With respect to the level of difficulty condition, the data was divided into four sets, one for each automation condition. The data was averaged across the two trials at the same level of difficulty and across the two raters. The resulting 24 data sets of 20 observations each were subjected to a two-tailed Wilcoxon Signed Ranks test.

8.3. Workload Measures

The overall NASA-TLX score or weighted workload measure was computed based on the rankings and ratings of the various workload dimensions. Four different data sets of 160 observations each were created for each of the team roles and workload types including: (1) AC rating of taskwork, (2) AC rating of teamwork, (3) IO rating of taskwork, and (4) IO rating of teamwork. Each of these data sets was subjected to the ANOVA models described previously. The trial order effect was significant for the AC weighted workload scores of both *taskwork* and *teamwork;* therefore, the full ANOVA model was used in those analyses. The order effect was not significant for either of the IO weighted workload scores, so the reduced model was used in those analyses. The order effect indicated that the ACs perceived higher workload for the trial order condition in which the high task difficulty was presented first (high, low, low, high) as compared to the order condition in which the low task difficulty condition was presented first (low, high, high, low). Because the radar display presented to the AC shows the number of targets in a spatial display, the presentation of 8 targets in the first trial, as opposed to the second trial (following more experience with the low task difficulty condition), may have affected the workload ratings of the AC more than the IO. Thus, analyses of workload ratings for the AC include order in the model while the analyses of workload ratings for the IO do not.

The data set of IO ratings of team workload was non-normal, based on significance of the Shapiro-Wilks test statistic. Transformations of the data set failed to correct for this non-normality. Neter et al. (1990) state that non-normality of data without accompanying non-constant variance generally leads to shifts or adjustments in the alpha level of the F-test. For example, an alpha level of 0.05 may actually represent a probability in the range of 0.04 to 0.065. Evaluation of the residual plots for this data set indicated that the variance was constant. Therefore, the final analysis was conducted using the ANOVA. Because of the non-normality, the more rigorous significance level of $p < 0.01$, commonly used in human factors research, was applied to this data set. *p*-values in the range of 0.01 to 0.05 were not considered significant.

9. RESULTS

9.1. Subject Characteristics and Team Makeup

The 40 teams included 17 male-male teams, 14 male-female teams, and 9 female-female teams. Team gender makeup was partially balanced across automation conditions (with a difference of no more than 1 of each gender combination between automation conditions). The average age of the AC was 26.5 years and the average age of the IO was 24.6 years. Eight teams in each automation condition were made up of team members who knew each other while two teams in each automation condition were made up of team members who did not know each other. Within the 32 teams who knew each other, 10 teams indicated that they were very close (a spouse or roommate), 15 teams indicated that they were close friends or worked together, and 7 teams indicated they were friends or colleagues. With respect to PC experience, the average response (on a scale of 1 = none to 5 = frequent) of the AC was 4.6 and the average response of the IO was 4.8. With respect to video game experience, the average response of the AC was 3.2 and the average response of the IO was 3.1.

9.2. Team Effectiveness Measures

9.2.1. Level of Difficulty Effects

Table 15 summarizes the means and standard deviations of each of the team effectiveness measures by level of difficulty. The ANOVAs on four of the six effectiveness measures revealed significant effects due to task difficulty. For target points ($F(1, 36) = 27$, $p < 0.0001$) and damage points ($F(1,32) = 81$, $p < 0.0001$), there were a greater number of points scored in the high difficulty condition than in the low difficulty condition. This was consistent with expectations given that there were more targets in the high difficulty condition, providing greater scoring opportunity for both target points and damage points. With respect to the total score ratio ($F(1, 36) = 7.1$, $p < 0.05$), there was a higher ratio of points scored to targets presented in the low difficulty condition than in the high difficulty condition. This indicates that at a low level of difficulty, teams were more successful on a per target basis in classifying and resolving targets correctly. With respect to the damage ratio ($F(1,32) = 21$, $p < 0.0001$), there was a lower ratio of damage points to targets presented for the low difficulty condition than for the high difficulty condition. This indicates that teams were better able to manage the pace of the task and not allow enemy targets to pass through missile defenses in the low difficulty condition as compared to the high difficulty condition. Figure 10 displays the average team effectiveness ratio scores by level of difficulty.

Table 15. Mean effectiveness scores by level of difficulty.

	Target Points**	Target Ratio	Damage Points**	Damage Ratio**	Total Points	Total Score Ratio*
Low	2187 (636)	12.9 (3.5)	741 (232)	4.4 (1.5)	1447 (759)	8.5 (4.4)
High	2734 (855)	12.4 (3.0)	1164 (357)	5.5 (2.0)	1569 (1039)	6.9 (4.3)

(Note: Standard deviations for each measure and difficulty condition are presented in parenthesis. An asterisk (*) indicates a significant effect at $p < 0.05$. Two asterisks (**) indicate a significant effect at $p < 0.01$.)

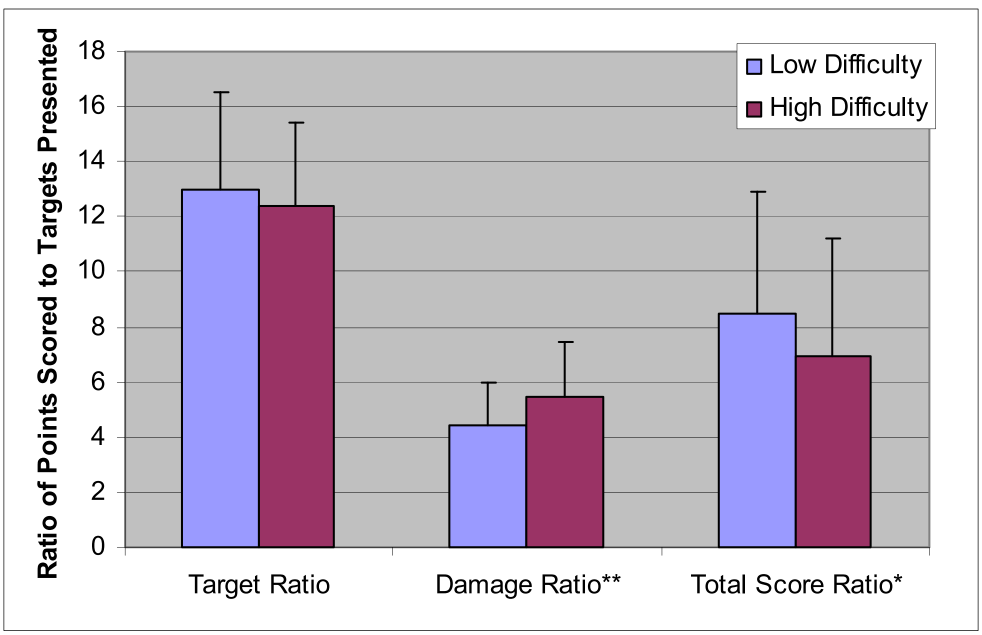

Figure 10. Average target, damage, and total score ratio by level of difficulty.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect at $p < 0.05$. Two asterisks (**) indicate a significant effect at $p < 0.01$.)

9.2.2. Automation Condition Effects

With respect to automation condition, none of the ANOVAs on team effectiveness measures yielded a significant result. Figure 11 displays the average target points, damage points, and total points scored for the four automation conditions. Although the effect was not significant, the DS condition had the highest mean target points and the lowest mean damage points (resulting in the highest total score). The DS condition also appeared to have a slightly higher standard deviation with respect to target points than the other three conditions.

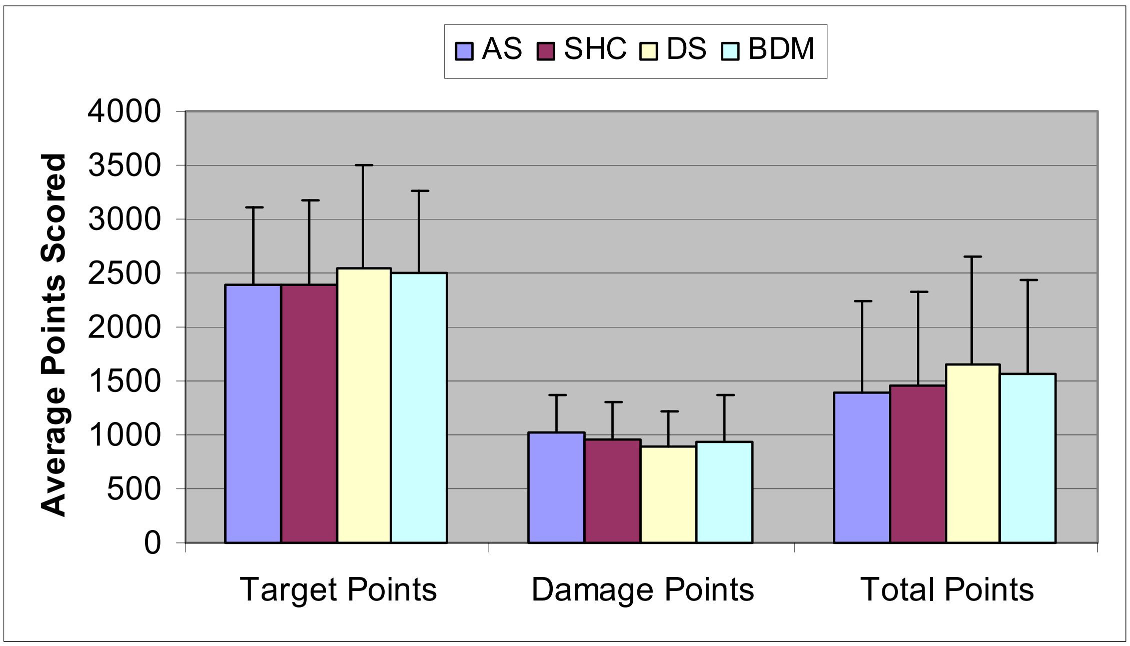

Figure 11. Average target, damage, and total points scored by automation condition. (Note: Error bars represent + 1 standard deviation above the mean.)

9.2.3. Automation Condition and Level of Difficulty Interactions

There was a significant interaction between the automation condition and level of difficulty in terms of the ratio of damage points to targets presented ($F(3,32) = 3.94$, $p < 0.05$). Figure 12 graphically presents the interaction effect of the automation condition and level of task difficulty on the damage ratio.

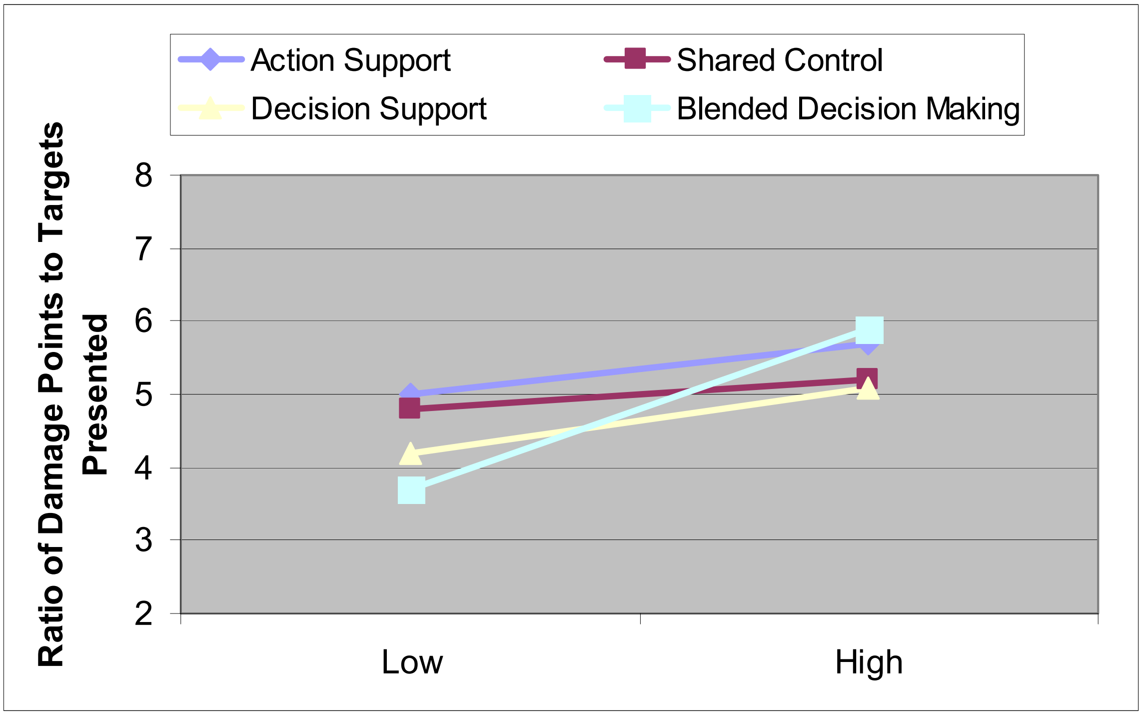

Figure 12. Damage point ratio by automation condition and level of difficulty.
(Note: The scale of damage ratio represents approximately plus and minus 1 standard deviation from the highest and lowest condition means.)

The results of Duncan's Multiple Range test on the different levels of the automation condition and task difficulty interaction are shown in Table 16. The conditions are sorted in the table from lowest to highest damage ratio, so that the best performance is indicated at the top of the table. The results of this analysis indicated that none of the automation conditions were significantly different from one another under the high task difficulty condition. However, under low task difficulty, the BDM condition lead to significantly better performance in terms of the damage ratio than the SHC and AS conditions. It is also interesting to note that the BDM condition led to the best performance in the low difficulty condition (significantly better than two of the other three conditions), but the worst performance in the high difficulty condition (not significantly different from the others).

Table 16. Results of Duncan Multiple Range test on the automation condition and level of difficulty interaction for damage ratio.

Duncan Grouping				Mean Damage Ratio	Automation Condition	Level of Difficulty
A				3.7	BDM	Low
A	B			4.2	DS	Low
	B	C		4.8	SHC	Low
	B	C	D	5.0	AS	Low
	B	C	D	5.1	DS	High
	B	C	D	5.2	SHC	High
		C	D	5.7	AS	High
			D	5.9	BDM	High

(Note: Means with the same letter to the left are not significantly different from one another at alpha <0.05.)

9.3. Team Coordination Measures

9.3.1. Communication Counts

Table 17 presents the results of the correlations analyses on the ratings of the two raters for all seven types of communications that were tallied. Significant correlations ($p < 0.0001$) were obtained for all seven communication types. The correlation coefficients for all communication types were greater than 0.9 with the exception of *other communications*.

Table 17. Inter-rater correlations for counts of specific communication types.

Communication Type	Pearson Correlation Coefficient, r (N=160)	Prob > H_0: $\rho = 0$, P
Classification request from AC with response from IO	0.98	< 0.0001
Classification request from AC without response from IO	0.98	< 0.0001
Unsolicited classification from the IO	0.99	< 0.0001
Reliability request from the IO with a response from the AC	0.93	< 0.0001
Unsolicited reliability report by the AC	0.96	< 0.0001
Other communications	0.79	< 0.0001
Total communications	0.96	< 0.0001

Figure 13 shows the average number of communications for the four automation conditions and two levels of difficulty. With respect to the automation condition, the

Kruskal-Wallis test was not significant for either the low difficulty or high difficulty data sets. However, given the reduced power of the test mentioned previously, it is worth noting that there was a trend toward significance for both the low difficulty (T(3) = 5.8, $p < 0.15$) and high difficulty (T(3) = 6.7, $p < 0.10$) data sets, most likely due to the lower communications in the SHC condition compared to other conditions.

There was no difference in the total number of communications for the two levels of difficulty.

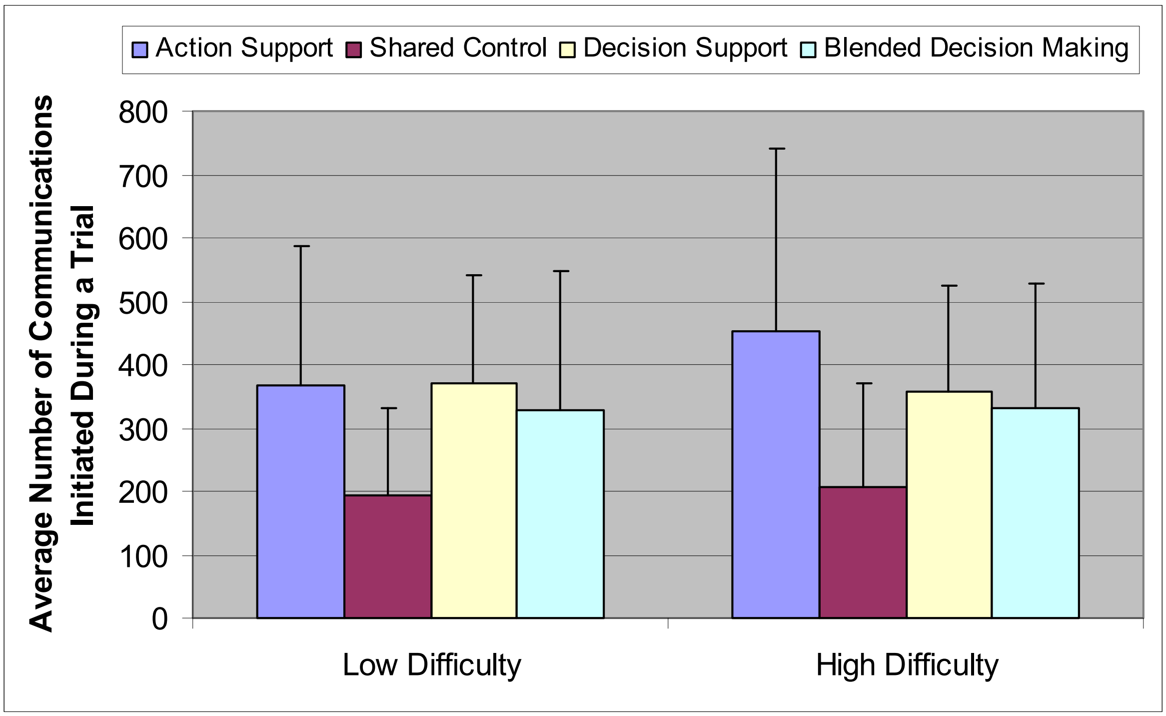

Figure 13. Average number of communications by automation condition and level of difficulty. (Note: Error bars represent + 1 standard deviation above the mean.)

The Kruskal-Wallis tests on each of the communication types indicated significant effects of automation condition for the following communication types and levels of difficulty (see Figures 14 and 15):

- Reliability request with response, low difficulty, T(3) = 11.4, $p < 0.01$,
- Unsolicited reliability reports, low difficulty, T(3) = 15.8, $p < 0.01$, and
- Unsolicited reliability reports, high difficulty, T(3) = 14.4, $p < 0.01$.

For these three significant effects, the Kruskal-Wallis test was repeated on pairs of the automation conditions. Based on these comparisons, the following effects were observed:

- For reliability request with a response under the low task difficulty condition, the SHC condition had fewer communications than both the DS condition ($T(1) = 12.2$, $p < 0.01$) and the BDM condition ($T(1) = 7.1$, $p < 0.01$).
- For unsolicited reliability reports under the low task difficulty condition, the SHC condition had significantly fewer reports than the AS ($T(1) = 8.7$, $p < 0.01$), DS ($T(1) = 12.1$, $p < 0.01$), and BDM ($T(1) = 8.7$, $p < 0.01$) conditions.
- For unsolicited reliability reports under the high task difficulty condition, the SHC condition had significantly fewer reports than the AS ($T(1) = 8.3$, $p < 0.01$), DS ($T(1) = 12.1$, $p < 0.01$), and BDM ($T(1) = 6.1$, $p < 0.05$) conditions.

All of the significant effects of automation on the communication counts appear to be due to the reduced need for communication of reliability information from the AC to the IO under the SHC condition, because the reliability information is included on the IO's display.

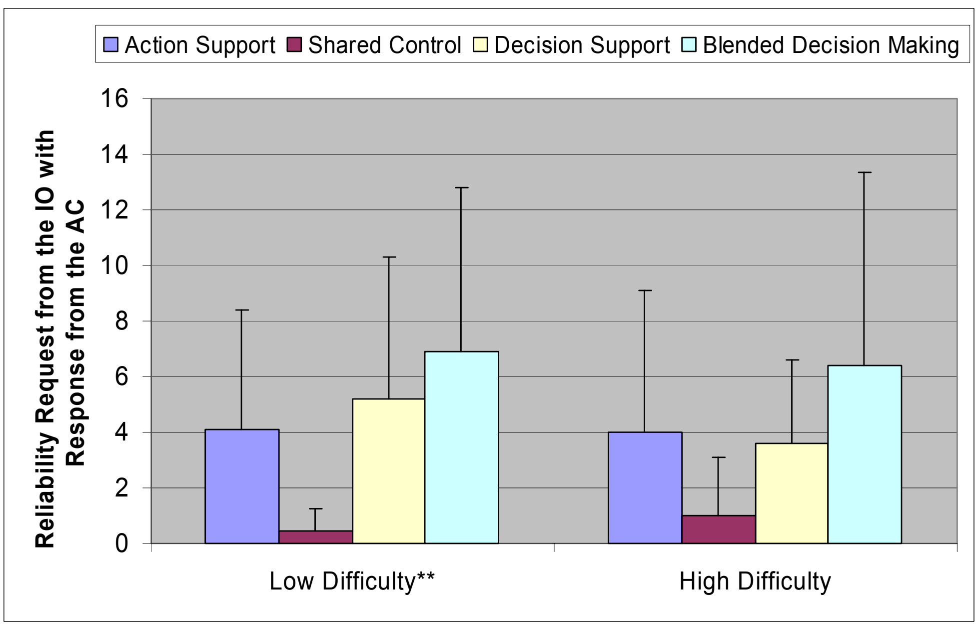

Figure 14. Average number of reliability requests with a response by automation condition.
(Note: Error bars represent + 1 standard deviation above the mean. Significant effects of automation on the specific types of communication for at least one of the low or high difficulty data sets are annotated with an asterisk (*) for $p < 0.05$ and two asterisks (**) for $p < 0.01$.)

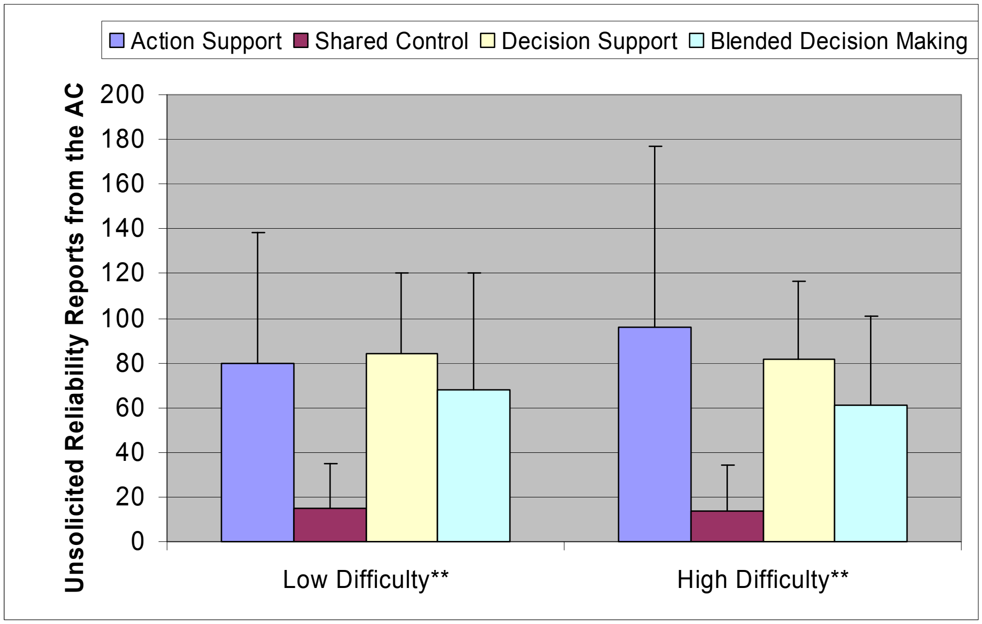

Figure 15. Average number of unsolicited reliability reports by automation condition.
(Note: Error bars represent + 1 standard deviation above the mean. Significant effects of automation on the specific types of communication for at least one of the low or high difficulty data sets are annotated with an asterisk (*) for $p < 0.05$ and two asterisks (**) for $p < 0.01$.)

As mentioned in the data analysis section, the power of the statistical test for detecting differences between automation conditions for several of the communication types was low. For these communication types (Figure 16), the following trends were observed:

- Classification request without response, high difficulty, T(3) = 6.9, $p < 0.10$,
- Unsolicited classifications, low difficulty, T(3) = 6.1, $p < 0.15$,
- Unsolicited classifications, high difficulty, T(3) = 5.9, $p < 0.15$,
- Reliability request with response, high difficulty, T(3) = 7.5, $p < 0.10$, and
- Other communications, high difficulty, T(3) = 6.7, $p < 0.10$.

Recall that *classification request without a response* does not imply that the request from the AC was ignored by the IO, instead it indicates that the IO did not respond verbally and very likely responded through the user interface with color coding of a target. For the trend on classification request without response, this may indicate that teams in the AS condition were more likely to receive classification requests and/or were more likely to use the interface to respond compared to teams exposed to the SHC condition. The trend for unsolicited classifications may indicate that subjects in the SHC and DS conditions were more likely to follow a strategy of the IO choosing the order in which to classify targets (and thus providing unsolicited classifications), as compared to teams in the AS condition (where the AC was more likely to request target classifications). The trend for *other communications* suggests that teams in the decision support condition verbalized more *other communications* than teams exposed to the BDM condition.

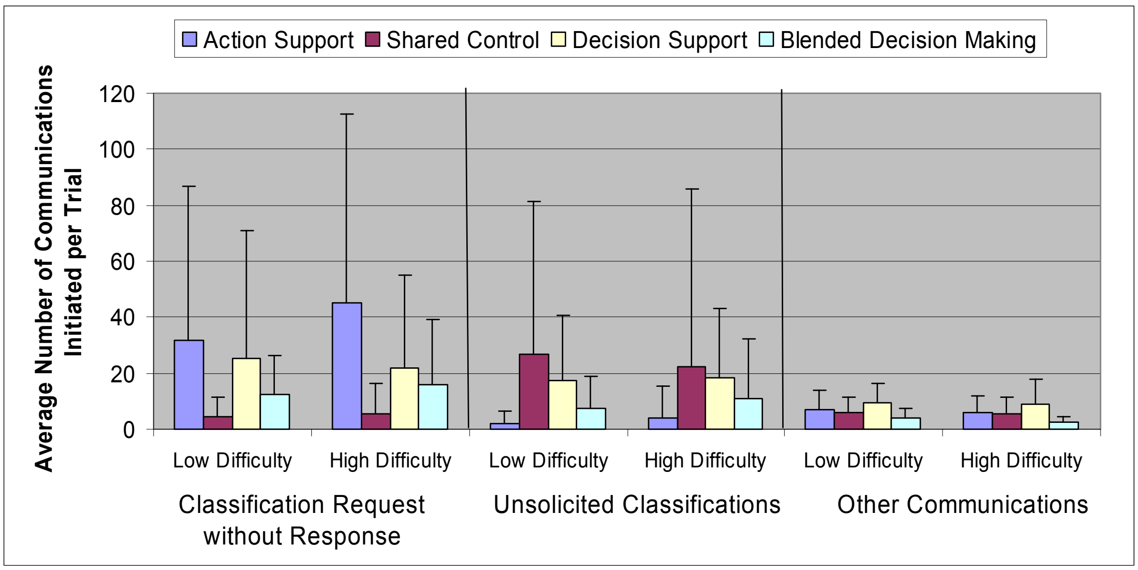

Figure 16. Average number of communications of a specific type by automation condition. (Note: Error bars represent + 1 standard deviation above the mean.)

With respect to level of difficulty, the Wilcoxon Signed Ranks tests indicated a significant effect for only one of the seven communication types. In the BDM data set, there were significantly more *other communications* under the low difficulty condition than in the high difficulty condition ($Z = 2.28$, $p < 0.05$). Figure 17 displays the average number of *other communications* by level of difficulty for the four automation conditions.

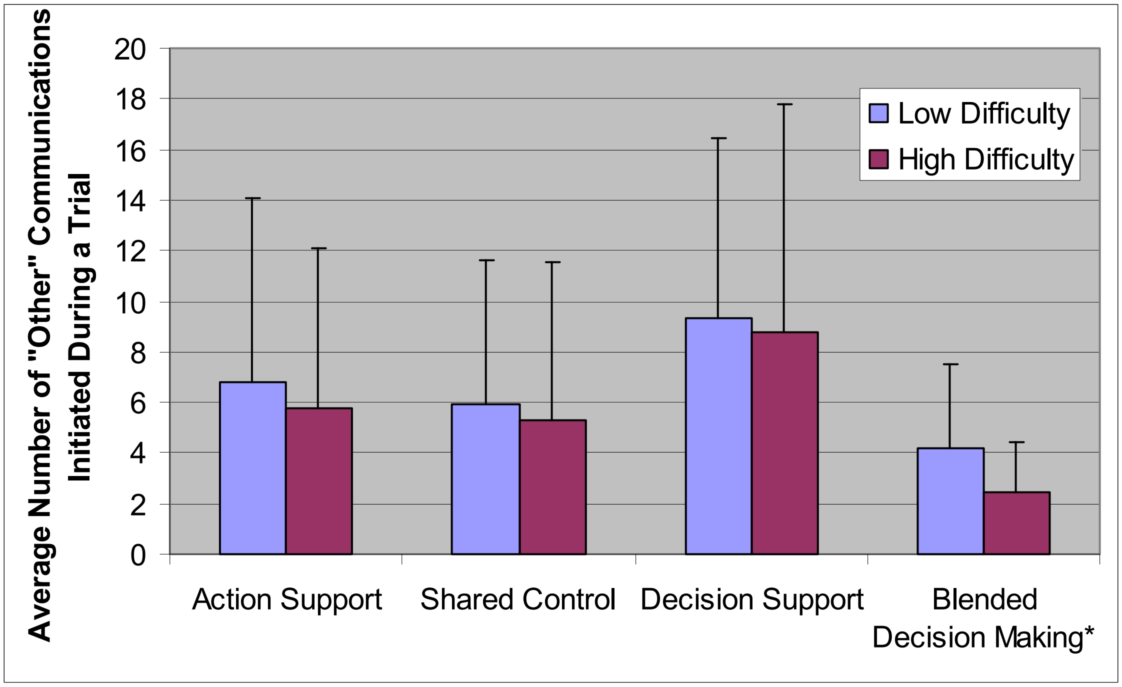

Figure 17. Average number of "other communications" by automation condition and level of difficulty.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect of level of difficulty at $p < 0.05$.)

9.3.2. Anticipation Ratio

Figure 18 presents the mean anticipation ratio for the four automation conditions. A higher anticipation ratio indicates that more information was transferred compared to the information that was requested. The Kruskal-Wallis test revealed a significant effect of automation under both the low ($T(3) = 9.2$, $p < 0.05$) and high difficulty conditions ($T(3) = 9.2$, $p < 0.05$). The comparison of the four automation conditions revealed that the for the low difficulty condition, SHC had a higher anticipation ratio than the AS condition ($T(1) = 6.2$, $p < 0.05$), the DS condition ($T(1) = 5.7$, $p < 0.05$), and the BDM condition ($T(1) = 5.9$, $p < 0.05$). For the high difficulty condition, SHC had a higher anticipation ratio than the AS condition ($T(1) = 5.1$, $p < 0.05$) and the DS condition ($T(1) = 7.0$, $p < 0.01$).

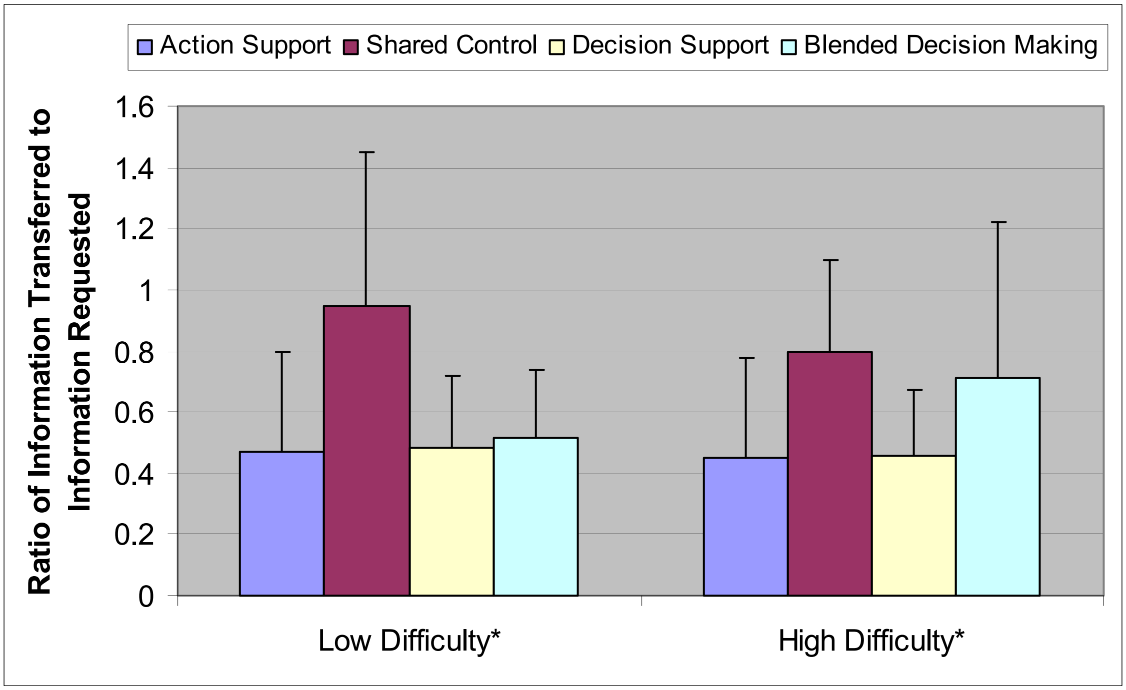

Figure 18. Anticipation ratio by automation condition.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect of automation condition at $p < 0.05$.)

9.3.3. Correlations Between Communication Counts and Team Effectiveness Measures

The correlation analyses of the communication counts and team effectiveness measures revealed significant results for *classification requests with a response*, *unsolicited classifications*, and *other communications*. The correlation coefficients and *p*-values for each of the six team effectiveness measures and these three communication types are shown in Table 18. There were no significant correlations for any of the other communication types or for the anticipation ratio. Since lower damage points and damage ratio indicate better performance, a negative correlation of these team effectiveness measures with a type of communication indicates that the communication was associated with higher performance. The results from the table indicate that *classification request with response* tended to be associated with lower performance (negative correlation with target points, total points and total score ratio, and positive correlation with damage ratio). *Unsolicited classification* was associated with better performance measured in terms of the damage ratio. *Other communications* was associated with better performance in terms of total points.

Table 18. Correlations of three communication types and the six team effectiveness measures.

Communication Type	Pearson Correlation Coefficient, *r,* N=160 Prob > H_0: ρ = 0, P					
	Target Points	Target Ratio	Damage Points	Damage Ratio	Total Points	Total Score Ratio
Classification request from AC with response from IO	-0.27** < 0.0005	-0.15 < 0.06	0.14 <0.09	0.27** <0.0005	-0.29** <0.0005	-0.23** <0.005
Unsolicited classification from the IO	0.006 < 0.94	-0.08 < 0.4	-0.11 < 0.16	- 0.17* < 0.04	-0.08 < 0.33	0.01 < 0.90
Other communications	0.14 < 0.08	0.13 < 0.10	-0.09 < 0.25	-0.11 < 0.15	0.16* < 0.05	0.14 <0.07

(Note: An asterisk (*) indicates a significant effect at $p < 0.05$. Two asterisks (**) indicate a significant effect at $p < 0.01$.)

Correlations of the communication types with each other also revealed a number of significant relationships. The results are shown in Table 19. Not surprisingly, the most common communication types (classification requests with and without a response, unsolicited classifications and unsolicited reliability reports) were positively correlated with total communications. More interesting, however, is the fact that *other communications* were negatively correlated with *total communications* which indicates that individuals who talked less in general were more likely to make other statements that might include leadership statements, situation awareness statements, and decision-making statements.

Also of interest is the fact that *classification request with response* correlated negatively with *other communications*. This indicates that teams that were more likely to use a strategy of the AC calling out specific targets for classification were less likely to make other statements. Since a greater number of *other communications* and fewer *classification requests with response* were associated with better performance and the two measures were also negatively correlated with each other, this suggests that all three measures were interrelated.

Table 19. Correlations between communication types.

Communication Type	Pearson Correlation Coefficient, *r*, N=160 Prob > H_0: ρ = 0, P					
	Classification request with response	Classification request without response	Unsolicited classification	Reliability request with response	Unsolicited reliability report	Other communica-tions
Classification request without response	-0.32** < 0.0001					
Unsolicited classification	0.016 < 0.9	-0.13 < 0.09				
Reliability request with response	-0.13 < 0.2	0.01 < 0.9	0.19* < 0.05			
Unsolicited reliability report	0.25** < 0.005	0.46* < 0.0001	-0.19* < 0.05	-0.10 < 0.3		
Other communications	-0.35** < 0.0001	-0.04 < 0.6	-0.14 < 0.08	0.02 < 0.8	-0.06731 < 0.4	
Total communications	0.62** < 0.0001	0.38** < 0.0001	0.18* < 0.05	-0.005 < 0.95	0.79** < 0.0001	-0.28** < 0.001

(Note: An asterisk (*) indicates a significant effect at $p < 0.05$. Two asterisks (**) indicate a significant effect at $p < 0.01$.)

The correlations of types of communications also show that *classification requests with response* were negatively correlated with *classification requests without a response*. This indicates that teams tended to either follow a strategy of the IO responding verbally to a request (from the AC) or the IO responding using the interface instead of a combination of the two. In addition, *unsolicited reliability reports* were positively correlated with both of the classification request counts. A number of teams followed a strategy of the AC calling out a target number and the highest reliability sensor in sequence. This correlation reflects this type of strategy whether the IO responded verbally or through the interface.

Finally, *reliability request with response* was correlated positively with *unsolicited classifications*. This suggests that for teams where the IO tended to verbally report classifications without a request from the AC, the IO was also more likely to request reliability reports from the AC.

9.3.4. Team Coordination Ratings

Table 20 presents the correlations of the ratings of the two raters on the five dimensions of teamwork and total teamwork. The correlation coefficients were all highly significant and ranged from $r = 0.30$ to $r = 0.56$. The coefficients were slightly lower than was expected based on initial pilot testing and the results of previous studies using a similar team coordination rating method (Brannick et al., 1993; Travilian et al., 1993; Volpe et al., 1996). Based on these studies, teamwork dimensions with inter-rater reliabilities greater than 0.45 were accepted for further analysis. These included leadership, situation assessment, and total teamwork. In the present research, the inter-rater reliabilities for these dimensions are also low. However, the correlations were all highly significant (indicating that a substantial portion of the variance in the ratings of Rater 1 can be explained by the ratings of Rater 2 and vice-versa) and, since the data for the two raters was averaged in the analyses, any differences between the raters only serve to make the results more conservative.

Table 20. Inter-rater correlations for ratings of team coordination.

Teamwork Dimension	Pearson Correlation Coefficient, r (N=160)	Prob > H_0: $\rho = 0$, P
Leadership	0.49	< 0.0001
Communication	0.32	< 0.0001
Decision Making	0.30	< 0.0001
Assertiveness	0.40	< 0.0001
Situation Assessment	0.54	< 0.0001
Total	0.56	< 0.0001

Figure 19 presents the average total rating of teamwork (inter-rater reliability = 0.56) for each of the four automation conditions under the two levels of difficulty. Higher ratings indicate higher teamwork skills. Figure 20 presents the average total ratings for those teamwork dimensions for which the inter-rater reliabilities were greater than 0.45 including leadership and situation assessment, under each of the four automation conditions. The Kruskal-Wallis test on leadership, situation assessment, and total teamwork (with the data separated by low and high difficulty conditions) indicated the following significant effects of automation:

- Leadership, under high difficulty (T(3) = 7.9, $p < 0.05$);
- Situation assessment, under low difficulty (T(3) = 8.1, $p < 0.05$); and
- Total teamwork rating, under both low difficulty (T(3) = 8.9, $p < 0.05$) and high difficulty (T(3) = 9.8, $p < 0.05$).

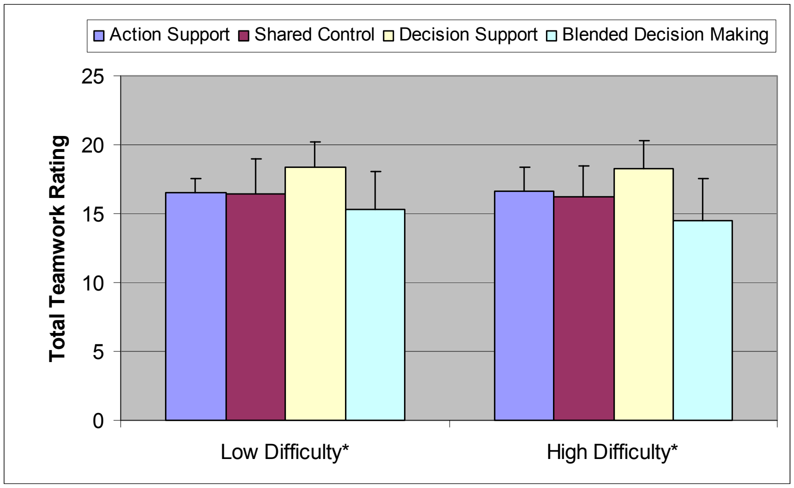

Figure 19. Average rating of total teamwork across automation conditions.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect of automation condition at $p < 0.05$.)

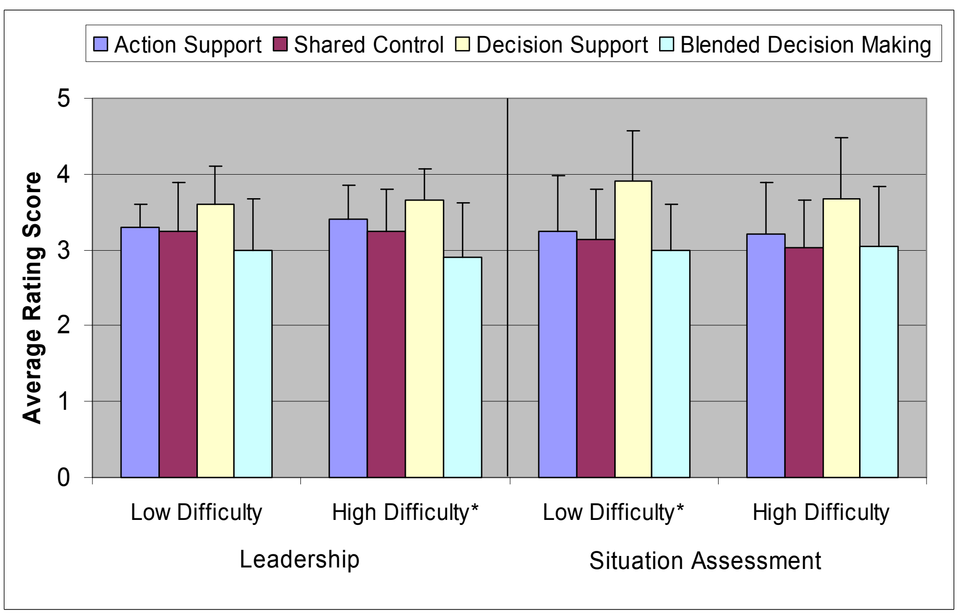

Figure 20. Average rating of leadership and situation assessment across automation conditions.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect of automation condition at $p < 0.05$.)

For each of the significant effects of automation, additional Kruskal-Wallis tests were conducted to compare each pair of automation conditions and determine which conditions were significantly different from each other. These post-hoc analyses yielded the following results:

- For leadership under the high task difficulty setting, the DS condition received higher ratings than the BDM condition ($T(1) = 6.5$, $p < 0.05$).
- For situation assessment under the low task difficulty setting, the DS condition received higher ratings than SHC ($T(1) = 5.1$, $p < 0.05$) and BDM ($T(1) = 6.9$, $p < 0.01$).
- For total teamwork under the low task difficulty setting, the DS condition received higher ratings than AS ($T(1) = 5.0$, $p < 0.05$) and BDM ($T(1) = 7.2$, $p < 0.01$).
- For total teamwork under the high task difficulty setting, the DS condition received higher ratings than BDM ($T(1) = 8.1$, $p < 0.01$).

There were no significant differences on any teamwork dimensions between the AS, SHC, or BDM automation conditions.

There were no significant differences in team coordination ratings due to level of difficulty for any of the automation conditions.

9.4. Workload Measures

9.4.1. Level of Difficulty and Automation Condition Effects

Figure 21 presents the weighted workload scores for each of the team member roles and workload types by level of difficulty. All four workload measures revealed significant effects due to level of difficulty. For AC taskwork ($F(1,32) = 19.1$, $p < 0.0001$), AC teamwork ($F(1,32) = 9.4$, $p < 0.005$), IO taskwork ($F(1,36) = 63.9$, $p < 0.0001$), and IO teamwork ($F(1,36) = 47.7$, $p < 0.0001$), the rating of workload was higher under the high difficulty condition than the low difficulty condition.

There were no significant main effects of automation on workload ratings for either the IO or the AC on either taskwork or teamwork.

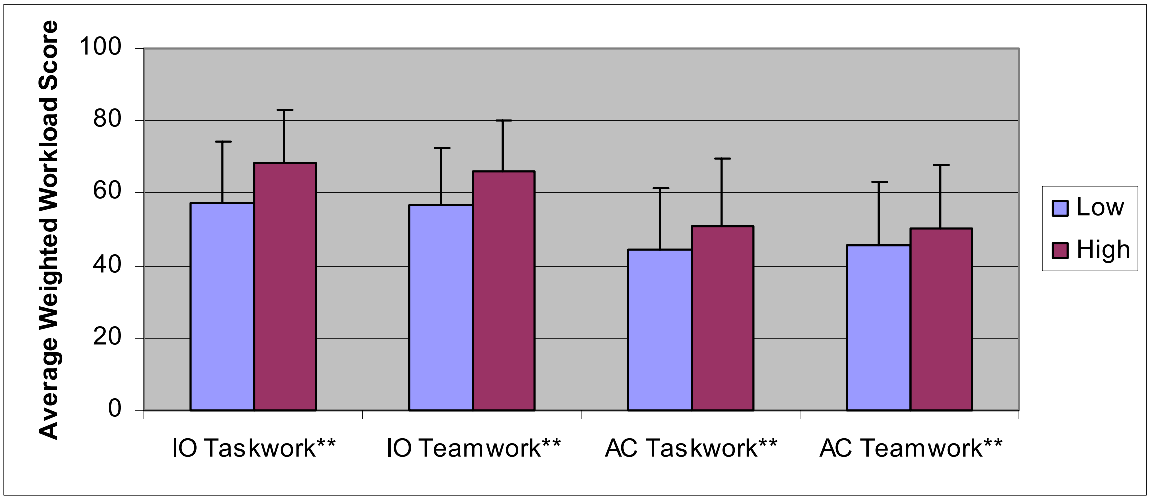

Figure 21. Average weighted workload score by level of difficulty for each of the workload measures.
(Note: Error bars represent + 1 standard deviation above the mean. An asterisk (*) indicates a significant effect at $p < 0.05$. Two asterisks (**) indicate a significant effect at $p < 0.01$.)

9.4.2. Automation Condition and Level of Difficulty Interactions

There was a significant interaction of automation condition and level of difficulty on the IO ratings of taskwork ($F(3,36) = 3.5$, $p < 0.05$)). A similar effect was revealed for teamwork ($F(3,36) = 3.1$, $p < 0.05$), however, this effect is not considered significant due to the stricter requirements for this non-normal data set (as described in the data analysis section). Consequently, only the interaction effect for taskwork is presented in detail here. Figure 22 shows the automation condition by level of difficulty interaction for the IO's rating of taskwork.

The results of Duncan's Multiple Range test on the levels of automation and the level of difficulty interaction are shown in Table 21. The mean workload scores are sorted in ascending order so that conditions with the lowest ratings of workload are listed at the top. This analysis revealed that under the high level of difficulty, the BDM condition received higher workload ratings than all other conditions. Under the low level of difficulty, BDM also received higher workload ratings than all other conditions. In addition, under the low level of difficulty, the DS condition received higher workload ratings than the AS and SHC conditions. Another interesting observation is that under the low level of difficulty, the BDM condition was not rated significantly different than all the other automation conditions under the high level of difficulty.

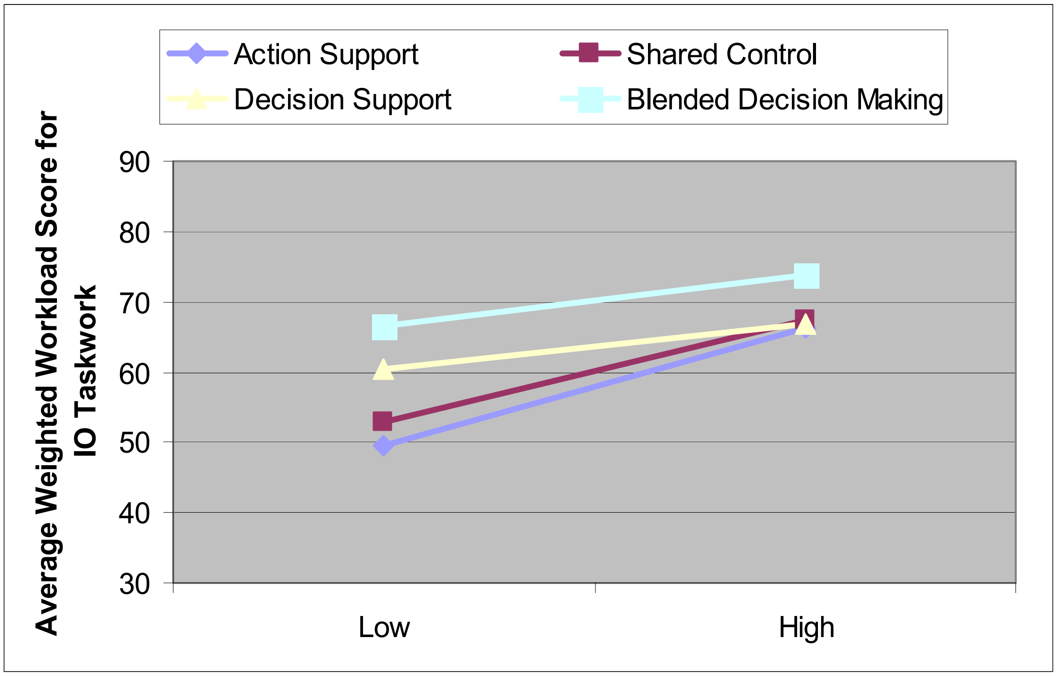

Figure 22. Automation condition by level of difficulty interaction for IO rating of taskwork. (Note: The scale of weighted workload score represents approximately ± 1 standard deviation from the highest and lowest condition means.)

Table 21. Results of Duncan Multiple Range test on the interaction of automation condition and level of difficulty interaction for IO taskwork.

Duncan Grouping				Mean Weighted Workload Score	Automation Condition	Level of Difficulty
A				50	AS	Low
A				53	SHC	Low
	B			60	DS	Low
		C		66	AS	High
		C		67	BDM	Low
		C		67	DS	High
		C		68	SHC	High
			D	74	BDM	High

(Note: Means with the same letter to the left are not significantly different from one another at $\alpha < 0.05$.)

10. DISCUSSION

10.1. Level of Automation Effects

10.1.1. Team Effectiveness Measures

With respect to the team effectiveness measures, it was hypothesized (see Table 12, H1) that performance would be poorest in the AS condition and best (Table 12, H2) in the BDM condition. The results indicated no significant main effects due to automation condition. However, for the BDM condition performance was higher than the other conditions with respect to damage ratio under the low level of difficulty only. On average, the AS condition had the lowest mean team effectiveness scores and the DS condition produced the highest scores; however, the difference between them was not significant.

These results suggest that advanced forms of automation providing information analysis and organization capabilities (DS and BDM) may have offered some performance advantage over the other low-level automation conditions in terms of preventing enemy aircraft from reaching the home base (negative events). In particular, under the low task difficulty condition, the combination of the prioritized list of aircraft and the automated classification information (BDM condition) allowed teams to reduce the damage ratio as compared to the low-level automation conditions. However, under high levels of task difficulty, this performance advantage was not present which may have been due to increased workload associated with considering the additional information in that condition.

10.1.2. Communication Counts

With respect to communication counts, it was hypothesized that there would be fewer overall communications in the SHC condition than in the other conditions (Table 12, H3). The SHC condition did exhibit the lowest mean number of total communications, with 100-200 fewer exchanges than the other three conditions. However, this effect was not statistically significant.

It was also hypothesized (Table 12, H4) that there would be lower communication counts related to reliability for the SHC condition than the other automation conditions. This proved to be true with SHC producing fewer unsolicited reliability reports than all other conditions. In addition, under the low level of difficulty, there were fewer reliability requests from the IO in the SHC condition as compared to the DS and BDM conditions. These results suggest that the addition of reliability information on the IO display in the SHC condition served to reduce communications between the AC and IO regarding sensor reliability.

Although there were no other significant effects on communication counts due to the automation condition, there were some trends that may support different communication patterns for the different automation conditions. The SHC condition appeared to have fewer classification requests from the AC and more unsolicited classifications from the IO. This suggests that in this condition, the IO was more likely to choose the order of processing targets with the AC only calling out specific targets when necessary. The DS condition also appeared more likely to follow this type of communication pattern as compared to the BDM and AS conditions.

In addition, there was a trend to suggest a greater number of *other communications* in the DS condition compared to the BDM condition. Communications tallied in the *other communications* category tended to include the following:

- statements related to situation assessment, such as reports of the current score or the state of the targets (e.g., "all friendly");
- leadership statements like "good job"; and
- decision-making statements and discussions.

Anecdotally, teams that developed more advanced strategies generally had a greater number of tallies of *other communications*. Advanced strategies included consideration of information from all the sensors and sometimes target responses to specific missile types; that is, if the correct missile did not kill the target, then the classification must have been wrong.

There were no specific hypotheses regarding effects on anticipation ratio by automation condition. Results revealed a higher anticipation ratio in the SHC condition compared to all other automation conditions. This suggests that teams in the SHC condition provided more information in their communications compared to the number of requests for information. There are two possible reasons for this. First, teams in the SHC condition tended to follow a pattern of the IO choosing the order in which to address targets rather than the AC calling out each target, which reduced the number of information requests. Second, since the IO had reliability information displayed on his or her screen, the IO would simply look at this information rather than request it from the AC.

10.1.3. Team Coordination Ratings

Team coordination ratings assessed the frequency and quality of specific teamwork behaviors including communication, leadership, decision-making, situation assessment, and assertiveness. Because of low inter-rater reliabilities, only the dimensions of leadership and situation assessment, and the total teamwork measure were analyzed. The SHC condition was hypothesized to have the highest ratings for communication, leadership, decision-making and total teamwork (Table 12, H5). The results did not support this hypothesis. There was no evidence to indicate higher teamwork ratings of any kind for the SHC condition compared to other conditions. In fact, the DS condition had the highest teamwork ratings in leadership (under high task difficulty only), situation assessment (under low task difficulty only) and total teamwork. While the SHC condition reduced the need for information transfer between the team members, it did not appear to foster more teamwork with respect to *other communications* that would have provided evidence to the raters of skill in decision-making or leadership.

There are two possible reasons why the teamwork ratings for the SHC condition were not higher than the other conditions, even though team members had more shared information available. One is that the SHC condition allowed team members to simply work independently and there was less interaction between the team members. Therefore, they did not develop the teamwork skills of communication, leadership, and decision-making. A second explanation is that the limited number of communications between team members simply provided less data for the raters to make assessments of teamwork. The rating method requires observations of specific behaviors for the teams to receive higher than average ratings.

Results partially supported the hypothesis that ratings of situation assessment would be lower for SHC compared to the other conditions (Table 12, H6). The SHC condition was rated lower in situation assessment than the DS condition under low task difficulty. This provides support for the assertion that the additional information as part of the SHC displays for both the IO and the AC reduced the need for communications related to situation assessment.

The team coordination results supported the hypothesis that automation of the decision selection aspect of the task (the BDM condition) would result in lower team coordination ratings (Table 12, H7). On average, the BDM condition received the lowest teamwork ratings across all dimensions. The BDM condition was rated significantly lower than the DS condition in leadership, situation assessment, and total teamwork. One explanation for the lower teamwork ratings in the BDM condition is that the additional information provided by automation of the decision selection aspect of the task increased the workload for the IO, requiring him or her to consider the information (particularly when it disagreed with the IO's own conclusions). This often meant that the IO took more time to respond to the AC for a particular target. This resulted in impatience on the part of the AC, who may have reacted by calling out targets that were nearing the center. The IO may then have experienced additional workload in responding to the AC's request. This sometimes caused the IO to abandon the target he or she was currently working on to search for the target called out by the AC. This type of interaction left little time for the team to exhibit behaviors that represented good decision-making, leadership, or situation assessment.

The results on team coordination did not support the hypothesis that ratings would be higher in the AS condition as compared to the DS condition (Table 12, H8). In fact, the opposite was true for ratings of total teamwork. The higher ratings of teamwork for the DS condition, compared to the other three conditions, coincide with the trend for a higher number of *other communications* under the DS condition. Since *other communications* included statements of situation assessment, leadership, and decision-making, it follows that teamwork ratings of these characteristics would be higher for teams that exhibit a higher number of *other communications*.

One possible explanation for this is that the DS condition provided team members with more time to make these types of statements. It was observed during test trials that *other communications* occurred more during periods when all of the aircraft on the radarscope had been classified as friendly. The prioritized listing of targets in the DS condition may have allowed teams to classify and shoot down enemies first, leaving them with more frequent periods of "friendly skies".

Because of low inter-rater reliabilities regarding the dimension of communications, this aspect of teamwork could not be adequately assessed. Anecdotally, the automation conditions did not appear to influence the use of standard terminology in any way.

10.1.4. Workload Ratings

Results did not support the hypotheses that workload would be lower for the more highly automated conditions (Table 12, H9 and H10). In fact, IO ratings of taskwork and teamwork were highest in the BDM condition. For the IO rating of taskwork, workload was higher in the BDM condition than all other conditions. Under the low level of task difficulty, IO ratings of taskwork were higher in the DS condition than in the AS and SHC conditions. This result suggests that, for the IO, the automated prioritization of targets and automation of the decision aspect of the task caused an increase in workload. With respect to the automated target sort, the interface changed frequently in order to update the list of targets. This introduced performance problems for the IO since these changes could occur while he or she was observing information on a particular target or at the time he or she was trying to press a classification button for the target. With respect to automation of the decision selection aspect of the task, the IOs seemed to treat this as additional information to consider in making a decision, rather than simply using the automation to reduce the decision-making load. Even though teams were told that the automation was highly reliable (as good as any human could be) except when the sensors had equal reliability, few if any of the teams appeared to adopt a strategy of using only the automation when any one sensor exhibited higher reliability than another. The increased workload for the DS and BDM conditions suggest that any performance advantage of the automated prioritization of targets and decision selection aspect of the task came at a cost of higher workload for the IO.

The workload results did not support the hypotheses that teamwork ratings would be greater for the AS condition compared to the SHC condition (Table 12, H11) and that taskwork ratings would be greater for the SHC condition compared to AS condition (Table 12, H12). In fact, the ratings of taskwork and teamwork didn't appear to adequately distinguish task workload from team workload. Throughout all of the automation and difficulty conditions for both the IO and the AC, ratings of taskwork and teamwork were very similar, indicating that team members were not making distinctions between the two concepts in their ratings.

10.2. Task Difficulty Effects

It was hypothesized that there would be higher target points and damage points for the high difficulty condition compared to the low difficulty condition (Table 12, H13). The greater number of targets in the high difficulty condition resulted in higher points scored on both measures. While there was a higher target ratio for the low difficulty condition compared to the high difficulty condition (Table 12, H14), the difference was not significant. With respect to the damage (Table 12, H15) and total score ratios, the difference between the low difficulty condition and high difficulty condition was significant. The lower damage ratio indicates that teams were better able to manage the pace of the task in the low difficulty condition. The higher total score ratio provides some evidence that teams performed better in terms of decision-making in the low difficulty condition as compared to the high difficulty condition.

As hypothesized, the workload ratings were higher in the high difficulty condition than in the low difficulty condition (Table 12, H16). This was true across all four measures of workload.

Contrary to expectation, the team coordination ratings were not higher in the low difficulty condition than in the high difficulty condition (Table 12, H17). The team coordination rating measure did not appear to be sensitive to task difficulty manipulations.

Finally, there were no differences in the total number of communications across high and low difficulty conditions. However, there was a difference in the number of *other communications* between the low and high difficulty conditions for the BDM automation condition. Teams had fewer *other communications* in the high difficulty condition than in the low difficulty condition. Under high difficulty teams had less time available for communications other than the necessary target classifications and reliability information.

10.3. Automation Condition and Task Difficulty Interactions

In general, the results did not support the hypothesis that level of automation effects on team effectiveness and team coordination would be more pronounced in the high difficulty condition than in the low difficulty condition (Table 12, H18). In fact, some automation condition effects were more pronounced in the low difficulty condition than in the high difficulty condition. The performance advantage of the BDM condition was apparent only in the low difficulty condition. This coincides with a higher number of *other communications* under BDM in the low difficulty condition as compared to the high difficulty condition. In addition, differences in workload ratings for the DS condition compared to the AS and SHC condition were apparent only in the low difficulty condition. The low difficulty condition provided teams with more time to consider the additional information provided by the automation, as evidenced by more *other communications*. Since teams in the DS and BDM conditions took advantage of the additional time in the low difficulty condition to consider the information provided by the automation, there were greater differences in workload across automation conditions at low difficulty. Teams in the AS and SHC condition may have viewed this time as a chance to take a break.

10.4. Team Strategies and General Characteristics of High Performing Teams

This section considers the effects of automation on both the strategies of teams and characteristics of high performance teams. Of particular interest is whether certain of the automation conditions promote (or discourage) characteristics or strategies of high performing teams. First, ideal strategies with respect to the TDT are presented and the influences of automation conditions on the selection or development of those strategies during the experimental trials are discussed. Second, the results of the experiment are discussed with reference to previous research that identifies specific characteristics of high performing teams.

10.4.1. Theatre Defense Task Strategies

The results of the correlations between communication counts and team performance, anecdotal observations during the data collection, and the results of the team coordination ratings were used to develop a model of the ideal strategy for the TDT. A common strategy undertaken by teams was for the AC to call out or request target classifications in order of the proximity of the targets to the home base. The AC might have provided a "best sensor" report at the same time (e.g., "26 A"). The IO would then classify the target and report either verbally or through the interface what the target type was. The IO's classification may have been preceded by a request for reliability information. This strategy was used across all four automation conditions. There is some empirical evidence to suggest that this strategy was used more for the AS and BDM conditions as compared to the SHC and DS conditions.

On the basis of the experimental data, the following four steps represent the best task strategy:

1. The IO chooses the order of processing except when targets are close to the center of the display or there is an impending collision. This reduces the need for the AC to report target numbers and for the IO to visually search for a specific target number in the list. This is even more beneficial in the case of the DS and BDM conditions because of the re-sorting of the target list that occurred when an aircraft was cleared.
2. The AC gives frequent unsolicited sensor reliability updates. If the IO selects the order of processing targets, he or she is usually able to keep pace with the task, such that targets are classified fairly soon after they enter the display. This is important because general reliability reports are accurate for all targets as they enter the screen. For some targets that have traveled close to the center of the display (or moved slowly), the AC may want to give sensor positions for that

target based on where the sensors were when that target entered screen. For optimal decision-making, updates from the AC should include actual percentages or reliability levels (e.g., “low”, “medium”, or “high”) and not simply the identification of the “most reliable” sensor.

3. The IO uses the interface to send target classifications. Using color codes makes it easy for the AC to quickly find and kill enemy targets. It also provides a reminder for those that have already been identified as friendly. However, using the interface to color code targets is less critical than Numbers 1 and 2 above. Since the AC has a lower workload, he or she generally has time to visually search for enemies and is able to remember which targets have already been identified as friendly. Therefore, verbal target classifications would not negatively affect performance to a great degree. In addition, it is not especially important that the IO be specific in his or her classifications. Simply selecting any “enemy” or “friendly” button generally results in good team performance.
4. In the BDM condition, the best strategy is for the IO to simply follow the automated classification, except when all reliabilities are equal.

Because of the workload differential between the AC and the IO and because of the compelling nature of the radar display in terms of presenting target proximity to the center of the display, teams did not automatically make the strategy decision for the IO to determine the order of processing the targets. However, this appeared to be the most critical decision in terms of maximizing team effectiveness while minimizing workload. Since the order of target presentation on the IO display was generally different than a target proximity order that would be followed by the AC, there was sometimes a conflict between team members regarding the order of processing targets.

Figure 23 presents a model showing the relationship between the decision regarding which team member controls the order of target processing, team communication behaviors, automation condition, and team performance. Solid black linkages in the Figure are supported either by statistically significant correlations, significant ANOVA or non-parametric test results, or by trend data from the experiment. Trend evidence is included only for specific types of communication where the power of the statistical test was low. Gray lines link additional factors that may be related to the strategy decision based on anecdotal evidence during the experiment. The automation condition is thought to influence the decision for the IO to control the order of processing and team situation assessment and decision-making statements. This is based on trends suggesting that automation conditions influenced *classification requests* and *other communications* as well as significant differences in team coordination ratings across automation conditions. In particular, the SHC and DS conditions may have had fewer *classification requests* than the AS and BDM conditions. The DS condition may have had more *other communications* and received higher team coordination ratings than the other automation conditions.

Air Commander leadership is also thought to affect the decision for the IO to control processing because teams in which the AC displayed the leadership characteristic of sensitivity to requests and inputs by the IO were more likely to follow this strategy. In addition, teams with a more assertive IO, who clearly recognized and upheld a preference for following the order of target processing presented on their display appeared to be more likely to follow this strategy. There is also reason to believe that teams with an AC who had a better understanding of the IO's task were more likely to follow the strategy of the IO controlling processing order, since the AC would recognize that requesting specific targets placed an additional load on the IO.

If a team chose to follow the strategy of the IO choosing the order in which to process targets, they were likely to have fewer verbalizations and were likely to verbalize more situation assessment, decision-making, and leadership statements. In the case of situation assessment, leadership, and decision-making statements, it is unclear whether this behavior influenced the decision for the IO to control the order of processing or whether this behavior was a result of that decision. Teams that followed the strategy of the IO choosing the order to process targets had higher team effectiveness scores. Anecdotally, when the automation condition was DS or BDM, teams that followed this strategy tended to process enemies first, which led to higher team effectiveness scores. It was also observed that teams that followed this strategy had more time and developed more advanced decision-making strategies.

Finally, it was unclear whether the advanced decision-making strategies led to more situation assessment and decision-making statements or vice-versa.

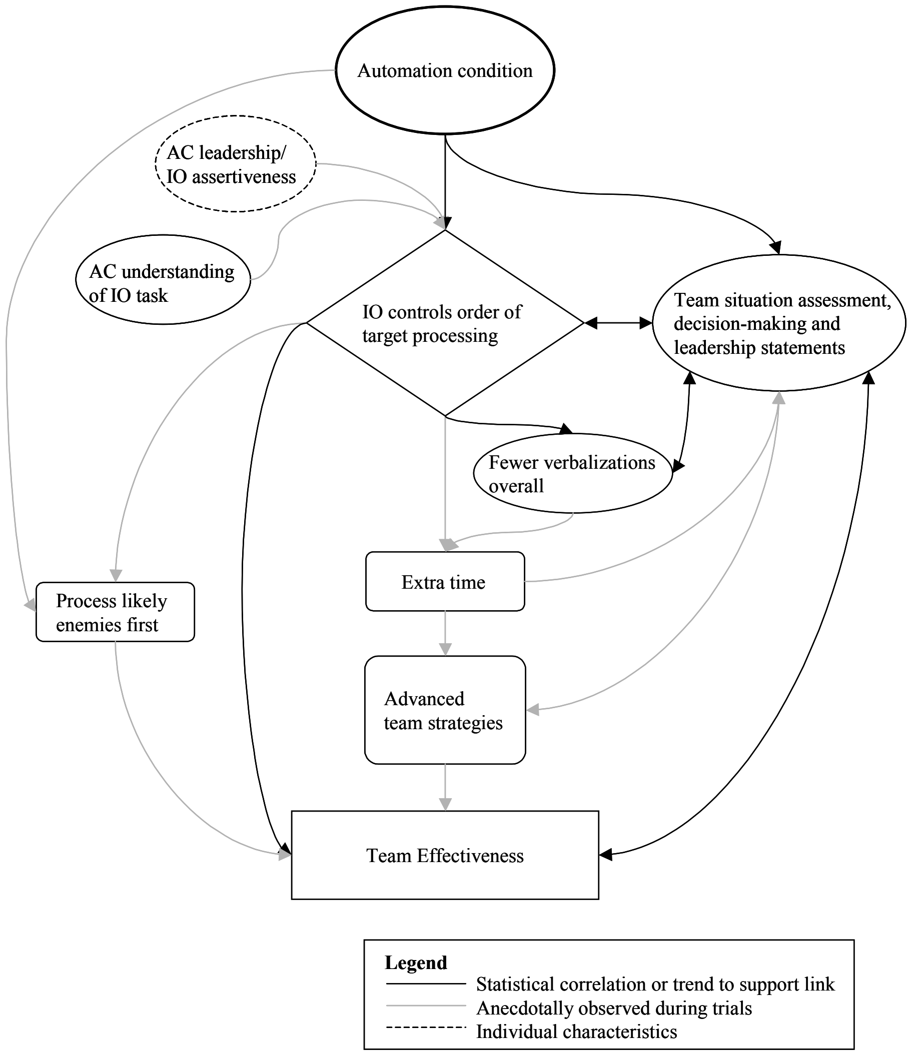

Figure 23. Relationship between decision to control order of processing and team effectiveness.

Considering the model in Figure 23, the trends toward reduced classification requests and increased unsolicited classifications for the SHC and DS condition, as well as the significant findings of a higher anticipation ratio for the SHC condition and higher team coordination ratings for the DS condition, we may conclude that automation of information acquisition and information analysis, within the context of the TDT, supports teamwork behaviors that ultimately lead to improved team effectiveness.

10.4.2. High Performing Teams

High performing teams exhibit efficient verbal communication, verbalizing only task relevant details and relying more on unsolicited reports rather than specific requests (Johannesen, Cook, and Woods, 1994; Urban et al., 1993). The correlations between types of communication and performance measures in the TDT support this observation. First, a greater number of classification requests (inefficient communication) were associated with poorer performance. In addition, a greater number of unsolicited classifications (efficient communication) were associated with better performance.

The results of this experiment revealed that the SHC condition had a higher *anticipation ratio* than the other conditions, suggesting more efficient and unsolicited communications. In addition, there were trends to suggest that teams in the SHC and DS conditions may have been more likely to adopt a strategy of more efficient and unsolicited communications with respect to target classifications. There were no differences due to automation condition with respect to the use of standard terminology, as evidenced by similar communication ratings across all automation conditions.

While the results support the inference that the shared information provided on the IO's and AC's display through the automation of information acquisition as part of the SHC condition supported more efficient communications, it is also possible that this information simply reduced the need for communications in general. Because of the high potential for error associated with team communications (Billings and Reynard, 1984; Helmreich and Foushee, 1993), one might conclude that automation of information acquisition is beneficial because it may reduce the need for communications. However, the reduced communications did not lead to higher team coordination ratings or any performance benefits. This is probably because of the additional visual load that was placed on the IO by presenting the information directly on his or her display. While the IO did not have to spend time requesting and listening to sensor reliability information, he or she did spend more time searching an already visually loaded display for needed information.

High performing teams also reduce the quantity of communication as workload increases (Jentsch et al., 1995; Urban et al., 1993; Wiener, 1993). None of the correlations between types of communication and performance measures provide a basis to determine whether high performing teams were more likely to reduce communication as workload increased. In the BDM condition, the number of *other communications* decreased with increasing workload. However, this probably does not reflect better performance on the part of teams in this condition. In fact, performance in the BDM condition was better than other conditions under the low difficulty condition but not in the high difficulty condition (where the reduction in communication occurred). Rather, this effect may be related to the higher overall workload associated with the BDM condition, requiring teams to severely limit anything other than the most critical task related communications in order to manage targets under the high difficulty condition. Considering Figure 17, the average number of *other communications* was lower in general for the BDM condition than for the other conditions (though not significantly). This supports the idea that the high level of workload for this condition did not allow time for *other communications*.

High performing teams conduct frequent planning and situation assessment sessions (Jentsch et al., 1995; Orasanu, 1990). Three of the measures in this experiment provided evidence of team planning and situation assessment behaviors. First, counts of the pre-session planning communications were intended to measure these behaviors. Unfortunately, observations of this type of communication were too few to provide any meaningful results. The ratings of situation assessment also provide evidence of team planning and situation assessment behaviors. Results of the situation assessment ratings indicated that situation assessment was rated higher for the DS condition than the BDM condition. A third measure that may indicate planning and situation assessment behaviors is the counts of *other communications*. The correlation of *other communications* with higher team effectiveness supports the assertion that this measure may be indicative of team situation assessment and planning behaviors. Like the ratings of situation assessment, this measure indicated a trend that teams in the DS condition might have had more *other communications* than teams in the BDM condition.

Both the ratings of situation assessment and the counts of *other communications* support the conclusion that the DS condition was more supportive of planning and situation assessment behaviors than the BDM condition. This may have been due to the difference in workload between the two conditions. However, workload was also lower for the AS and the SHC condition and these conditions did not exhibit better planning and situation assessment behaviors. In the case of the SHC condition, the limited interaction between team members may have carried on throughout all aspects of the task, including planning and situation assessment. With respect to the AS condition, the interaction between the team members appeared more likely to follow the pattern of the AC requesting a classification and the IO responding, which left little time for planning and situation assessment. In the DS condition, the prioritized listing of targets encouraged teams to follow a strategy of the IO determining the order in which to address the targets. This resulted in more efficient communications and allowed more time for planning and situation assessment and consequently a better team strategy and decision-making. Although the statistical analyses of the performance measures do not provide support for this reasoning, the DS condition did have the highest mean total points overall (see Figure 11). This condition also had the highest standard deviation of total points scored. This suggests that only some of the teams in the DS condition may have followed a strategy in which the IO determined the target processing order. This may be why there was no overall performance advantage for this condition.

A final characteristic of high performing teams is shared or common mental models of the task (Bolstad and Endsley, 1999a; Orasanu, 1990, 1993). In this experiment, team members in all automation conditions were cross-trained in the other team member role. They had an opportunity to observe the other team member's display and were given some (minimal) practice time in the opposite role to which they were assigned. In addition, team members were given ample opportunity to discuss the task and their strategies prior to performing the task. Team members were somewhat limited in this discussion in that they were required to use the radios for communications and were not able to see one another's displays at the time. There is no evidence to support differences between automation conditions in the development of shared or common mental models based on any of the pre-trial strategy discussions. The additional planning and situation assessment behaviors described above for the DS condition during the task may have resulted in more common mental models between team members over the course of the trials. However, there is no specific data to support this.

10.5. Comparisons to Previous Research on Automation and Teams

Previous studies comparing the performance of teams using automated systems with teams using more conventional systems have found few effects with respect to team effectiveness (Bowers et al., 1993; Wiener et al., 1991). Similarly, this experiment showed little difference across automation conditions in team effectiveness measures. However, automation of information analysis and decision selection had a positive effect on team performance under the low task difficulty condition, when the teams had time to consider all of the information provided by the automation. This improvement in team effectiveness came at a cost of increased workload.

With respect to quantity and quality of team communications, previous research has provided mixed results with automation associated with both increases and decreases in the amount of communication (Costley, Johnson, and Lawson, 1989; Wiener et al., 1991; Wise et al. 1992) and both improvements and decrements in the quality of communication (Bowers et al., 1993; Clothier, 1991; Costley, Johnson, and Lawson, 1989; Petridis et al., 1985; Straus and Cooper, 1989; Wiener et al., 1991). The preceding discussion of the results of this experiment provided some explanation for these contradictory findings. This study showed that the form of automation is important in influencing the quantity and quality of communications. Automation of information acquisition, in which the automated system provides more shared information to team members, may reduce the quantity of communications. Specifically, the SHC condition produced a higher *anticipation ratio* with team members providing more information in relation to the number of requests as compared to other forms of automation. Information analysis and decision selection automation may have little or no effect on the quantity of communications.

Automation of information analysis, in which the automated system analyzed the available information and reformatted it to provide the team with an aid in choosing an order in which to process targets was associated with higher quality communication. On the other hand, when the automation of information analysis was accompanied with automation of the decision selection aspect of the task, the advantages to team coordination were lost. The additional task complexity and resulting workload associated with considering the decision selection provided by the automated system negatively affected team coordination.

Bolstad and Endsley (2000) used a previous version of the TDT and compared the performance of teams in the AS and SHC conditions. They found that in a high difficulty condition, teams performed better in the SHC condition than in the AS condition. This finding was not upheld by this experiment. One main difference between the conditions employed by Bolstad and Endsley and those in the current study is that the reliabilities of the sensors in the current study were linked to the position of the AWACS aircraft. In contrast, in the Bolstad and Endsley study, the output of the sensors was pre-determined to present a number of specific combinations of matching, dissonant, or missing reports (even though teams were told that the information was linked to the position of the AWACS aircraft). Another important difference in the two studies is that for the current study, the training session was longer and more thorough (team members were cross-trained in both roles) and the trials were also longer. It may be possible that the effect reported by Bolstad and Endsley is only apparent for less experienced teams.

10.6. Team Workload

A secondary goal of this experiment was to evaluate the use of the NASA-TLX to separately measure task workload and team workload for both of the team members. While both measures of taskwork and teamwork were sensitive to differences in task difficulty and automation conditions and appeared to be sensitive to differences between the two team member roles, there was nothing to indicate that these measures targeted different workload constructs. The ratings of teamwork and taskwork were very similar across subjects and experimental conditions. In general, it may be difficult for individuals to subjectively separate the load associated with parts of a task described as their job from parts of a task requiring coordination with a team member. However, the measurement of workload separately for the two team members does provide valuable information with respect to the workload associated with different task conditions. In addition, the measurement of workload separately for the two team members allows researchers or designers to identify differences in workload based on roles which can then be used to redesign the task to more evenly distribute the load between team members.

10.7. Guidelines for Automated Systems Design

Based on the results of this experiment, the following guidelines regarding the design of automated systems in team environments are offered.

1. Automate information acquisition so that team members have access to the same information.

   Automation of information acquisition reduced the need for communication between team members. One caveat to this recommendation is to ensure that the automation does not overload any one sensory input modality. Automation of information acquisition generally results in the presentation of more visual information which might outweigh the benefits of a reduction in communication requirements. The presentation of this information should be designed carefully so that there is not a resulting increase in visual search time. In addition, consider other input modalities for presenting information such as synthesized speech or tactile input.
2. Automate intermediate steps in the task by providing advanced analysis of information.

   Providing information analysis automation, such as prioritizing tasks, can lead to improved team coordination and more advanced team strategies. If possible, present the advanced information analysis to both team members. As in the case of automation of information acquisition, the presentation of the information should consider both the visual load on system operators and any possible side effects, such as problems induced by information updates.
3. Consider automation of decision selection only in high risk, low time stress applications.

Automated decision selection can be helpful when there is enough time for the team to consider the decision selected by the automation, as it relates to system information. In cases where the automation is not highly reliable and there is time pressure for the team to act (or confirm the decision of the automation), automation of decision selection may be detrimental.

4. Balance workload between team members.

   In designing automated systems for teams, the workload for the team members should be assessed and tasks should be distributed so that workload is balanced.

11. CONCLUSIONS

11.1. The Effects of Automation on Teamwork

The goal of this experiment was to assess the effect of different forms of automation on the performance and coordination of teams. The results of the experiment have shown that different forms of automation have different effects on teamwork. Table 22 summarizes the different team performance findings based on the form of automation.

Increasing automation of information acquisition can lead to fewer team communications and a higher anticipation ratio; that is, teams transmit more information compared to the amount of information they request. This form of automation appeared to be more likely to lead to teams choosing a more effective strategy within the context of the TDT. However, these potentially positive effects on teamwork were not reflected in the team coordination ratings or team effectiveness measures. With respect to the team coordination ratings, the reduced communications between team members under information acquisition automation had the effect of reducing team coordination, even when it may have been beneficial. Secondly, the ratings of team coordination required observations of behaviors in the form of verbalizations and since there were fewer verbalizations in this automation condition, there were fewer observations on which to base ratings. More direct measures of team knowledge or team situation awareness (Cooke, Kiekel, and Bell, in press; Endsley, 1995; Entin and Entin, 2001) may provide insight into whether the automation of information acquisition actually leads to improved team situation awareness. With respect to the team effectiveness measures, it is possible that the increased information presented to the IO, which allowed for the reduction in communications, also added to the visual load and visual search time associated with the task; thus, negating any potential effects on outcome measures.

Increasing automation of information analysis resulted in higher team coordination ratings than conditions with less automation (the baseline condition), automation of earlier aspects of the task (information acquisition), and automation of later aspects of the task (automation of both information analysis and decision selection). These ratings appeared to be linked to a greater number of communications related to situation assessment, decision-making, and leadership and a greater likelihood to select a more efficient strategy within the context of the TDT. The information analysis automation allowed team members to process enemy targets first and may have provided team members with more time for communications that are representative of good teamwork. The higher workload ratings for this condition as compared to the baseline and information acquisition automation conditions under the low task difficulty condition suggest that teams were more likely to take advantage of any extra time in order to improve their performance rather than perceiving the time as a rest period. The team coordination benefits of this form of automation were not reflected in the team effectiveness measures.

Automation of the decision selection aspect of the task in addition to automation of information analysis resulted in higher team effectiveness measures under low task difficulty. Teams were able to take advantage of the additional information provided by the automation in order to improve their performance. However, this advantage came at the cost of higher workload. In the high task difficulty condition, there appeared to be a breakdown in this advantage. Teams no longer had better performance, their workload remained high, and they had no time for communications related to situation assessment, leadership, or decision-making.

Table 22. Summary of the effects of automation on team performance.

Form of Automation	**Findings**
Information acquisition	Decreased the quantity of communications Increased the anticipation ratio Led to the selection of a better task strategy
Information analysis	Increased ratings of team coordination Increased the quantity of communications related to situation assessment, decision-making, and leadership Led to the selection of a better task strategy Increased workload under low difficulty
Decision selection	Decreased ratings of team coordination Increased performance under low difficulty Increased workload

The results of this study were consistent with previous research in three main areas: (1) characteristics of high performing teams, (2) the effect of automation on teams, and (3) the influence of different levels of automation on human performance. High performing teams in this study exhibited characteristics identified by other studies to be indicative of high performing teams. These include efficient communications and frequent planning and situation assessment.

With respect to the existing body of research on teams and automation, the results have been mixed. The fact that differing forms of automation had different influences on team performance in this research aids in explaining conflicting historical findings. In general, researchers must be more specific in their descriptions and assessments of automation so that the results of research on team performance with complex systems may be generalized to other systems. The model of levels of automation by Parasuraman et al. (2000) is one way of specifying automation for this purpose. This research served to further the knowledge of the effect of automation on teams by exploring the types and levels of automation defined based on the Parasuraman et al. (2000) model and making specific reference to the taxonomy of LOAs developed by Endsley and Kaber (1999). The results represent a step forward in addressing the concerns of Wiener (1993) and Bowers (1996) that the effect of automation on both the quality and quantity of communications is not yet understood.

The results of this research are also consistent with previous research assessing the effects of different levels of automation on human performance. Previous work comparing the effect of various levels of automation on individuals has supported automation of psychomotor functions (e.g., automation of information acquisition and action implementation) and intermediate forms of automation (e.g., automation of information analysis) compared to higher-order cognitive functions such as decision selection (Endsley and Kaber, 1999; Kaber, Wright, and Clamann, 2002; Laois & Giannacourou, 1995; Ruff 2000). This experiment provides additional support in this area suggesting that automation of early and intermediate stages of processing may have benefits with respect to teamwork, while automation of decision selection may be more limited in the contexts in which it will be beneficial.

11.2. Caveats and Future Research Directions

One difficulty in making conclusions on a model of levels of automation, and attributing performance effects to the various stages of the model or degrees of automation, is that specific interface features intended to represent various automated functions may also influence operator performance and workload. Within the TDT, the SHC and BDM conditions both resulted in a greater visual load for the IO. Considering Multiple Resource Theory (Wickens, 1984), humans are better able to perform multiple tasks simultaneously when the inputs take advantage of different input channels (e.g., auditory and visual). Within teams, information can be received from a team member via the auditory channel at the same time the individual is receiving visual input. In the case of automation of information acquisition for this experiment, team coordination requirements were essentially "traded" for additional visual input for the IO. Thus, it becomes difficult to assess whether the results (or lack of results) were due to the degree and stage of automation of the task or due to changes in the operator input modalities, particularly because the IO's task was already visually intense. One way of dealing with this problem in future research would be to design tasks in a way that better maintains the balance of input modalities across conditions. (It may not be possible to perfectly balance the information content and information modality with different forms of automation.) For example, in the case of automation of information acquisition in the TDT, the automation could present target order and sensor reliabilities via synthesized speech output to the operator headsets.

One interesting effect of the automated prioritization of targets in a list for the IO was that it sometimes created a competition between the AC and the IO regarding the order in which to process targets. This was, in part, because only the IO was receiving the information analysis automation with respect to potential enemies. If the automated order were displayed to both team members, then this would result in better shared awareness and would possibly reduce the likelihood of this conflict. One way of presenting this automation to the AC without adding an additional interface feature to the AC's display would be to highlight potential enemy targets on the radarscope. This might promote following the order suggested by the automation and give a fairer depiction of the potential benefits of this level of automation.

Perhaps the most obvious limitation of this experiment was the lack of a more direct measure of team situation awareness or team knowledge. Future research in this area should include an additional dependent measure of team knowledge or situation awareness. A number of measures have been proposed in recent research. Cooke, Kiekal, and Bell (in press) have shown that measures of team knowledge assessed through questions related to taskwork knowledge (including "big picture" knowledge and knowledge of the other team member roles) immediately after training are predictive of team performance. The Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995) in which a task is frozen and probe questions are presented, as well as the mutual awareness measure (Entin and Entin, 2000) in which team members are asked to "think back" on specific events that focus on both the current task environment and the situation as it relates to other team member roles, may provide more direct indicators of team knowledge and situation assessment.

One final direction of future research in this area is to compare the effects of automation on decision-making strategies. For those teams that used the interface to classify targets, the existing data set contains all of the necessary information for comparing the output of the sensors, the reliability of each of the sensors, and the IO decisions regarding target classification. This data would allow for the determination of whether the different automation conditions affected the IO's decision to use more sophisticated decision-making strategies versus heuristics, such as always classifying a target based on the majority output of the sensors as enemy or friendly (and ignoring the sensor reliabilities). If the BDM condition, for example, does support more sophisticated decision-making in low difficulty conditions, it may be worthwhile to consider implementing this form of automation in conditions of high risk but low time stress.

12. REFERENCES

Andre, A. and Degani, A. (1997). Do you know what mode you're in? An analysis of mode error in everyday things. In Mouloua, M. and Koonce, J.M. (Eds.) Human Automation Interaction: Research and Practice. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bainbridge, L. (1987). Ironies of automation. In Rasmussen, J., Duncan, K., and Leplat, J. (eds.). New Technology and Human Error, pp. 271-283. New York: John Wiley and Sons.

Bell, H.H., and Lyon, D.R. (2000). Using observer ratings to assess situation awareness. In Endsley, M.R. and Garland, D.J. (eds.), Situation Awareness Analysis and Measurement, pp. 129 – 146. Mahwah, NJ: Lawrence Erlbaum Associates.

Bettenhausen, K.L. (1991). Five years of group research: What have we learned and what needs to be addressed? Journal of Management, 17, 345-381.

Billings, C.E. and Reynard, W.D. (1984). Human factors in aircraft incidents: Results of a 7-year study. Aviation Space and Environmental Medicine, October 1984.

Billings, C.E. (1997). Aviation Automation: The Search for a Human-Centered Approach. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bolstad, C.A. and Endsley, M.R. (1999a). Shared mental models and shared displays: an empirical evaluation of team performance. Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society.

Bolstad, C.A. and Endsley, M.R. (1999b). The effect of task load and shared displays on team situation awareness. Technical report SATECH-99-05. Marietta, GA: SA Technologies.

Bolstad C.A. and Endsley, M.R. (2000). The effect of task load and shared displays on team situation awareness. Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society.

Bowers, C.A., Braun, C.C. and Morgan, B.B. (1997). Team Workload: Its Meaning and Measurement. In Brannick, M.T., Salas, E. and Prince, C. (eds.). Team Performance Assessment and Measurement: Theory, Methods and Applications, pp. 85-108. Mahwah, N.J.: Lawrence Erlbaum Associates.

Bowers, C., Deaton, J., Oser, R., Prince C., & Kolb, M. (1993). The impact of automation on crew communication and performance. Proceedings of the Seventh International Aviation Psychology, pp. 758-761. Columbus: Ohio State University.

Bowers, C.A., Morgan, B.B., Salas, E., and Prince, C. (1993). Assessment of Coordination Demand for Aircrew Coordination and Training. Military Psychology, 5(2), 95-112.

Bowers, C.A., Oser, R.L., Salas, E. and Cannon-Bowers, J.A. (1996). Team performance in automated systems. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.243-263. New Jersey: Lawrence Erlbaum Associates.

Bowers, C., Salas, E., Prince, C. and Brannick, M. (1992). Games teams play: A method for investigating team coordination and performance. Behavior Research Methods, Instruments, and Computers, 24(4), 503-506.

Brannick, M.T., Roach, P.M., and Salas, E. (1993). Understanding team performance: A multimethod study. Human Performance, 6, 287-308.

Brannick, M.T. and Prince, C. (1997). An Overview of Team Performance Measurement. In Brannick, M.T., Salas, E. and Prince, C. (eds.). Team Performance Assessment and Measurement: Theory, Methods and Applications, pp. 3 -16. Mahwah, N.J.: Lawrence Erlbaum Associates.

Brannick, M.T., Prince, A., Prince, C., and Salas, E. (1995). The Measurement of Team Process. Human Factors, 37(3), 641-651.

Brehmer, B. (1987). Models of diagnostic judgments. In Rasmussen, J., Duncan, K., and Leplat, J. (eds.). New Technology and Human Error, pp. 87-95. New York: John Wiley and Sons.

Canon-Bowers, J.A., Tannenbaum, S.I., Salas, E., and Volpe, C.E. (1995). Defining competencies and establishing team training requirements. In Guzzo, R.A. and Salas, E. (Eds.) Team Effectiveness and Decision Making in Organizations. San Francisco: Jossey-Bass Publishers.

Card, S.K. (1989). Human factors and artificial intelligence. In Hancock, P.A., and Chignell, M.H (Eds.) Intelligent Interfaces: Theory, Research and Design. New York: Elsevier Science Publishers (North-Holland).

Chen, H., Houston, A., Nunamaker, J., and Yen, J. (1996). Toward intelligent meeting agents. Computer, 29(8), 62-69.

Clay, M.C. (1994). Subjective measures of mental workload for automatic processing tasks. Unpublished Masters Thesis. Raleigh, NC: North Carolina State University.

Clothier, C. (1991). Behavioral interactions across various aircraft types: Results of systematic observations of line operations and simulations. Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 332-337). Columbus: Ohio State University.

Comstock, Jr., J.R. and Arnegard, R.J. (1992). The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research. (NASA Technical Memorandum 104174). Hampton, Virginia; NASA Langley Research Center.

Cooke, N., Kiekel, P.A., and Bell, B. (in press). Addressing limitations in the measurement of team cognition. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.

Costley, J., Johnson, D., and Lawson, D., (1989). A comparison of cockpit communication B737-B757. Proceedings of the Fifth International Symposium on Aviation Psychology. (pp. 413-418). Columbus: Ohio State University.

Coury, B.G. and Semmel R.D. (1996). Supervisory control and the design of intelligent user interfaces. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 201-219. New Jersey: Lawrence Erlbaum Associates.

Dickinson, T.L. and McIntyre, R.M. (1997). A Conceptual Framework for Teamwork Measurement. In Brannick, M.T., Salas, E. and Prince, C. (eds.). Team Performance Assessment and Measurement: Theory, Methods and Applications. Mahwah, N.J.: Lawrence Erlbaum Associates.

Driskell, J.E. and Salas, E. (1992). Can you study real teams in contrived settings? The value of small group research to understanding teams. In Swezey, R.W. and Salas, E. (eds.) Teams Their Training and Performance, pp. 101-124/ New Jersey: Ablex Publishing Corporation.

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 97-101. Santa Monica, CA: Human Factors Society.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32 – 61.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. Human Factors, 37(1), 65-84.

Endsley, M.R. (1996). Automation and situation awareness. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.163-181. New Jersey: Lawrence Erlbaum Associates.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: A critical review. In Endsley, M.R. and Garland, D.J. (eds.), Situation Awareness Analysis and Measurement, pp. 3 – 32. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M.R. and Kaber, D.B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. Ergonomics, 42(3), pp. 462-492.

Endsley, M.R. and Kiris, E.O. (1995). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37(2), 381-394.

Fleishmann, E.A., and Zaccaro, S.J. (1992). Toward a taxonomic classification of team performance functions: Initial considerations, subsequent evaluations and current formulations. In Swezey, R.W. and Salas, E. (eds.), Teams: Their Training and Performance (pp. 31-56). Norwood, NJ: Ablex.

Foushee, H.C., Lauber, J.K., Baetge, M.M., & Acomb, D.B. (1986). Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations. (NASA Technical Memorandum 88322). Moffett Field, CA: NASA-Ames Research Center.

Foushee, H.C. & Helmreich R.L. (1988). Group interaction and flight crew performance. In E. L. Wiener & D.C. Nagel (Eds.), Human factors in aviation (pp. 189-227). San Diego: Academic Press.

Funk, K.H., Lyall, E.A. and Niemczyk, M.C. Flightdeck automation problems: Perceptions and reality. In Mouloua, M. and Koonce, J.M. (Eds.) Human Automation Interaction: Research and Practice. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hammer, J.M. and Small, R.L. (1995). An intelligent interface in an associate system. In Rouse, W.B. (Ed.), Human/Technology Interaction in Complex Systems, 7, 1-44. Greenwich, CT: JAI Press.

Harris, W.C., Hancock, P.A. and Arthur, E. (1993). The effect of taskload projection on automation use, performance and workload. In. R.S. Jensen & D. Neumeister (Eds.), Proceedings of the 7th International Symposium on Aviation Psychology (pp. 178-184 or 890a – 890f). Columbus: The Ohio State University.

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 & N. Meshkati (Eds.), Human Mental Workload (pp.139-178). North-Holland: Elsevier Science.

Helmreich, R.L. and Foushee, H.C. (1993). Why crew resource management?: The history and status of human factors training programs in aviation. In Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (1993). Cockpit Resource Management. San Diego: Academic Press, Inc.

Helmreich, R.L. and Schaefer, H., (1994). Team performance in the operating room. In Bogner, M.S. (Ed.) Human Error in Medicine. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Horrey, W.J. and Wickens, C.D. (2001). Supporting situation assessment through attention guidance, a cost-benefit and depth of processing analysis. Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, 316 – 320.

James, M., McClumpha, A., Green, R., Wilson, P., & Belyavin, A. (1991). Pilot attitudes to flight deck automation. Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 192-197). Columbus: Ohio State University.

Jentsch, F., G., Sellin-Wolters, S., Bowers, C.A., Salas, E. (1995). Crew coordination behaviors as predictors of problem detection and decision making times. Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, 1350 - 1353.

Johannesen, L.J., Cook, R.I., and Woods, D.D. (1994). Cooperative communications in dynamic fault management. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting, 225 - 229.

Kaber, D. B. & Endsley, M. R. (1997a). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, *16*(3), 126-131.

Kaber, D.B. & Endsley, M.R. (1997b). The combined effect of level of automation and adaptive automation on human performance with complex, dynamic control systems. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, 205 - 209.

Kaber, D.B. and Endsley, M.R. (1998). Team Situation Awareness for Process Control Safety and Performance. Process Safety Progress 17(1), 43 – 48.

Kaber, D. B. and Endsley, M. R. (in review). Level of automation and adaptive automation: Theory and empirical analysis. Submitted to Theoretical Issues in Ergonomics Science.

Kaber, D. B., Endsley, M. R., Wright, M. C. and Warren, H. L. (2002). The effects of levels of automation on performance, situation awareness, and workload in an advanced commercial aircraft flight simulation. (Final Rep.: NASA Langley Research Center Grant #NAG-1-01002). Hampton, VA: NASA Langley Research Center.

Kaber, D.B. and Riley, J.R. (1999). Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. International Journal of Cognitive Ergonomics, 3(3), 169-187. Lawrence Erlbaum Associates.

Kaber, D.B., Riley, J.M., Tan K.W., Endsley, M.R. (2001). On the design of adaptive automation for complex systems. International Journal of Cognitive Ergonomics, 5(1), 37 – 57. Lawrence Erlbaum Associates.

Kaber, D. B., Wright, M. C. and Clamann, M. P. (2002). Human response to adaptive automation of early sensory/information acquisition functions and later stages of information processing in complex system operations (Final Rep.: NASA Langley Research Center Grant #NAG-1-01039). Hampton, VA: NASA Langley Research Center.

Kanki, B.G. and Palmer, M.T. (1993). Communication and crew resource management. In Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (Eds.) Cockpit Resource Management. San Diego: Academic Press, Inc.

Kraiger, K. and Wenzel, L.H. (1997). Conceptual Development and Empirical Evaluation of Measures of Shared Mental Models as Indicators of Team Performance. In Brannick, M.T., Salas, E. and Prince, C. (eds.). Team Performance Assessment and Measurement: Theory, Methods and Applications, pp. 63 - 84. Mahwah, N.J.: Lawrence Erlbaum Associates.

Laois, L. and Giannacourou, M. (1995). Perceived effects of advanced ATC functions on human activities: Results of a survey on controllers and experts. In Proceedings of the Eighth International Symposium on Aviation Psychology (pp. 392-397). Columbus, OH: The Ohio State University.

Layton, C., Smith, P.J., and McCoy, C.E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. Human Factors, 36(1), 94-119.

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

Lester, J., Converse, S., Stone, B., Kahler, S. and Barlow, T. (1997). Animated pedagogical agent and problem-solving effectiveness: A large scale empirical evaluation. Proceedings of the Eighth World Conference on Artificial Intelligence in Education, pp. 23-30, Kobe, Japan, August 1997.

Lester, J., Voerman, J., Towns, S., and Callaway, C. (1997). Cosmo: A like-like animated pedagogical agent with deictic believability. Working Notes of the IJCAI '97 Workshop on Animated Interface Agents: Making Them Intelligent, pp. 61-69, Nagoya, Japan, August 1997.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zakland, A.L., Bittner, A.C., and Wherry , R.J. (1989). Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies (Technical Report 851). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

MacMillan, J., Entin, E.E., and Serfaty, D. (in press). From team structure to team performance: a framework. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.

Maes, P. (1994). Agents that reduce work and information overload. Communications of the ACM. 37(7), 31 - 40.

McIntyre, R.M., and Salas, E. (1995). Team performance in complex environments: What we have learned so far. In Guzzo, R. and Salas, E. (eds.), Team Effectiveness and Decision Making in Organizations (pp. 333-380). San Francisco: Jossey-Bass.

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

Morgan, B.B., Glickman, A.S., Woodard, E.A., Blaiwes, A.S., and Salas, E. (1986). Measurement of Team Behaviors in a Navy Environment (Tech. Rep. No. TR-86-014). Orlando, FL: Naval Training System Center.

Mosier, K.L., and Skitka, L.J., (1996). Human decision makers and automated decision aids: made for each other?. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.201-220. New Jersey: Lawrence Erlbaum Associates.

Muir, B. (1987). Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine Studies, 27, 527-539.

Neter, J., Wasserman, W., and Kutner, M.H. (1990). Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Boston, MA: Irwin.

Orasanu, J.M. (1990). Shared mental models and crew decision making. (Cognitive Science Laboratory Report #46). Princeton, NJ: Princeton University.

Orasanu, J.M. (1993). Decision-making in the cockpit. In Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (Eds.). Cockpit Resource Management. San Diego: Academic Press, Inc.

Orlady, H.W. (1989, October) Selected human factors in aviation. Paper presented at 6th Annual Meeting of the Mexican Association of Aerospace Medicine and the College of Professional Pilots of Mexico. Provencia Juriquilla, Mexico.

Parasuraman, R., Molloy, R., Mouloua, M., and Hilburn, B. (1996). Monitoring of automated systems. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.91-115. New Jersey: Lawrence Erlbaum Associates.

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

Parsons, H.M. (1985). Automation and the individual: Comprehensive and comparative views. Human Factors, 27, 99-111.

Prince, C., Chidester, T., Bowers, C.A., and Cannon-Bowers, J.A. (1992). Aircrew coordination training. In Swezey, R.W. and Salas, E. (eds.), Teams: Their Training and Performance (pp. 329-354). New York: Ablex.

Prince, C. and Salas, E. (2000). Team situation awareness, errors, and crew resource management: Research integration for training guidance. In Endsley, M.R. and Garland, D.J. (eds.), Situation Awareness Analysis and Measurement, pp. 3 – 32. Mahwah, NJ: Lawrence Erlbaum Associates.

Prinzel, L.J., Scerbo, M.W., Freeman, F.G. and Mikulka, P.J. A bio-cybernetic system for adaptive automation. Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, 1365 – 1369.

Riley, V. (1994). Human Use of Automation. Unpublished Doctoral Dissertation. Ann Arbor, MI: University of Minnesota.

Riley, V. (1996). Operator reliance on automation: Theory and data. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 19-35. New Jersey: Lawrence Erlbaum Associates.

Ruff, H., Draper, M. H. & Narayanan, S. (2000). The effect of automation level and decision aid fidelity on the control of multiple remotely operated vehicles. In D. B. Kaber and M. R. Endsley (Eds.), Human Performance, Situation Awareness and Automation: User-centered Design for the New Millenium (pp. 70-75). Madison, WI: Omnipress.

Ruffell Smith, H.P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. (NASA Technical Memorandum 78482). Moffett Field, CA: NASA-Ames Research Center.

Salas, E., Dickenson, T.L., Converse, S.A., and Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In Swezey, R.W. and Salas, E. (Eds.) Teams: Their Training and Performance. Norwood, New Jersey: Ablex Publishing Corporation.

Sarter, N. (1991). The flight management system: Pilots' interaction with cockpit automation. Proceedings o the International Conference on Systems, Man and Cybernetics, 1307-1310. Boston, MA: IEEE.

Sarter, N.B. (1994). Strong Silent, and "Out of the Loop": Properties of Advanced (Cockpit) Automation and their Impact on Human-Automation Interaction. Unpublished Doctoral Dissertation. Ohio State University.

Sarter, N.B. (1996). Cockpit automation: from quantity to quality, from individual pilot to multiple agents. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.117-136. New Jersey: Lawrence Erlbaum Associates.

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

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

Sarter, N.B. (2000). Multimodal communication in support of coordinative functions in human-machine teams. In Kaber, D.B., and Endsley, M.R. (Eds.), Human Performance, Situation Awareness, and Automation: User-centered Design for the New Millenium, pp. 47-50. Madison, WI: Omnipress.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 37-63. New Jersey: Lawrence Erlbaum Associates.

Scerbo, M.W. (2000). Adaptive automation: What's working and what's missing. Human Performance, Situation Awareness, and Automation: User-Centered Design for the New Millenium Conference Proceedings, p. 3.

Selcon, S.J. (1990). Decision support in the cockpit. Probably a good thing? Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomics Society (pp. 46-50). Santa Monica, CA: Human Factors and Ergonomics Society.

Serfaty, D., Entin, E.E., and Johnston, J. (1998). Team coordination training. In Cannon-Bowers, J.A. and Salas, E. (eds.) Making Decisions Under Stress, pp. 221-245. Washington, D.C.: American Psychological Association.

Serfaty, D. Entin, E.E., and Volpe, C. (1993). Adaptation to stress in team decision-making and coordination. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting. 1228-1232.

Sheridan, T. B. and Verplanck, W. L. (1978). Human and computer control of undersea teleoperators (Tech. Rep.). Cambridge, MA: MIT Man-Machine Laboratory.

Stout, R.J., Canon-Bowers, J.A., Salas, E., and Morgan, B.B. (1990). Does crew coordination behavior impact performance. Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting, 1382-1386.

Stout, R. and Salas, E. (1993). The role of planning in coordinated team decision making: Implications for training. Proceedings o-f the Human Factors and Ergonomics Society 37th Annual Meeting, 1238-1242.

Tesluk, P., Matthieu, J., Zaccaro, S., and Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In Brannick, M.T., Salas, E. and Prince, C. (eds.). Team Performance Assessment and Measurement: Theory, Methods and Applications. Mahwah, N.J.: Lawrence Erlbaum Associates.

Thornton, C., Braun, C., Bowers, C., and Moran, B. (1992). Automation effects in the cockpit: A low fidelity investigation. Proceedings of the Human Factors Society, 36th Annual Meeting, 1, pp. 30 – 34. Atlanta, GA: Human Factors Society.

Travillian, K.K., Volpe, C.E., Canon-Bowers, J.A., Salas, E. (1993). Cross-training highly interdependent teams: effects on team processes and team performance. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting 1243-1247.

Urban, J.M., Bowers, C.A., Monday, S.D. and Morgan, B.B. (1993). Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting 1233-1237.

Urban, J.M., Bowers, C.A., Monday, S.D. and Morgan, B.B. (1995). Communication for effective team performance. Military Psychology, 7, 123-129.

Veinott, E.S., and Irwin, C.M. (1993). Analysis of communication in the standard versus automated aircraft. Proceedings of the Seventh International Symposium on Aviation Psychology (584-588). Columbus: Ohio State University.

Volpe, C.E., Cannon-Bowers, J.A., Salas, E. and Spector, P.E. (1996). The impact of cross-training on team functioning: An empirical investigation. Human Factors 38(1), 87-100.

Weaver, J.L., Bowers, C.A., Salas, E. and Cannon-Bowers, J.A. (1995). Networked simulations: New paradigms for team performance and research. Behavior Research Methods, Instruments, and Computers, 27(1), 12-24.

Weaver, J.L., Morgan, B.B., and Hall, J. (1993). Team Decision Making in the Command Information Center: Development of a Low-Fidelity Team Decision Making Task for Assessing the Effects of Teamwork Stressors. (Technical Report submitted to the Office of Naval Research, Naval Training Systems Center, February, 1993).

Wickens, C.D. (1984). Engineering Psychology and Human Performance. Glenview, Illinois: Scott, Foresman and Company.

Wickens, C.D., Mavor, A.S., and McGee, J.P. (Eds.) (1997). Flight to the Future. Human Factors in Air Traffic Control. Washington, D.C.: National Academy Press.

Wickens, C.D., Mavor, A.S., Parasuraman, R. and McGee, J.P. (Eds.) (1998). The Future of Air Traffic Control: Human Operators and Automation. Washington, D.C.: National Academy Press.

Wickens, C. D. & Hollands, J. G. (2000). Engineering psychology and human performance (3rd ed.). New Jersey: Prentice Hall.

Wiener, E.L., Chidester, T.R., Kanki, B.G., Palmer, E.A. Curry, R.E. & Gregorich, S.E. (1991). The impact of cockpit automation on crew coordination and communication: I. Overview, LOFT evaluations, error severity, and questionnaire data. (NASA Contractor Report No. 177587). Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L. (1993). Crew coordination and training in the advanced-technology cockpit. In Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (Eds.) Cockpit Resource Management. San Diego: Academic Press, Inc.

Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (1993). Cockpit Resource Management. San Diego: Academic Press, Inc.

Wise, J.A., Guide, P.C., Abbott, D.W. & Ryan, L. (1992). Human factors in aviation safety: The effects of automation on corporate pilots (Interim Technical Report CAAR-15406-92-1). Daytona Beach, FL: Embry-Riddle Aeronautical University.

Woods, D.D. (1996). Decomposing automation: apparent simplicity, real complexity. In Parasuraman, R. and Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp.3-16. New Jersey: Lawrence Erlbaum Associates.

APPENDIX A – TEAM COORDINATION AND WORKLOAD RATING FORMS

Team Communication Count and Behavior Observation Form

Team # ______________
Trial # _______________

Auto Condition: AS SHC DS BDM
LOD: E D
Order Condition: 1 2

Event	Tally		Leadership		Communication		Decision-Making		Assertiveness		Situation Assess.	
			Good	Poor	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Target classification request from AC	With IO response	Without IO response										
Unsolicited target classification from IO												
Sensor reliability request from IO	With AC response	Without AC response										
Unsolicited sensor reliability update from AC												
Strategy suggestion	AC initiated	IO initiated										
Other	AC initiated	IO initiated										
Clarification	AC initiated	IO initiated										
Pre-trial strategy session	AC initiated	IO initiated										

Team # _____________
Trial # ______________

Auto Condition: AS SHC DS BDM
LOD: E D
Order Condition: 1 2

Comments:

Rating:
Leadership _____ **Decision-making _____** **Situation Assessment_______**
Communication ______ **Assertiveness________**

Observer Teamwork Rating – Definition Of Teamwork Dimensions And Behaviors

ASSERTIVENESS
Assertiveness refers to the willingness to make decisions, demonstrating initiative, and maintaining one's position until convinced otherwise by the facts.

Behaviors that suggest assertiveness include:
- Confronting ambiguities and conflicts
- Asking questions when uncertain
- Maintaining a position when challenged
- Making suggestions
- Stating an opinion on decisions, procedures, or strategies

DECISION-MAKING
Includes identifying possible solutions to problems, evaluating the consequences of each alternative, selecting the best alternative, and gathering information needed prior to arriving at a decision.

Behaviors that suggest decision-making skill include:
- Communicates possible solutions
- Gathers information to evaluate solutions
- Communicates consequences of alternatives
- Cross-checks information sources
- Selects the best alternative
- Development of plans
- Implements the decisions that were made

SITUATION ASSESSMENT
Situation assessment refers to the verbalization of information related to the perception of elements in the environment, the comprehension of their meaning in terms of task goals, and the projection of their status in the near future. Situation assessment verbalizations may serve to promote shared situation awareness between team members.

Behaviors that suggest situation awareness include:
- Situation assessment updates in which team members communicate the current state of the system
- Identification of problem situations and recognizing the need for action
- Exchange of information for the prevention of errors
- Noting deviations in SA between team members
- Demonstrated awareness (e.g., via verbal communication) of the ongoing mission status and the overall goal
- Integration of information from multiple sources
- Accurately prioritizing information and actions

LEADERSHIP
Team leadership involves providing direction, structure, and support for other team members. It does not necessarily refer to a single individual with formal authority over others. Team leadership can be shown by several team members.

Behaviors that suggest leadership skill include:
- Explains to the other team member exactly what is need from them during the task
- Listens to the concerns of the other team member
- Provides statements of team direction, strategy, or priorities for the task
- Sets goals for the team and orients the team toward those goals
- Provides feedback to the other team member regarding his/her performance

COMMUNICATION
Involves the exchange of information between two or more team members in the prescribed manner and by using proper terminology. One purpose of communication is to clarify or acknowledge the receipt of information.

Behaviors that suggest communication skill include:

- Verifies information prior to taking an action
- Acknowledges and repeats messages to ensure understanding
- Uses accurate terminology
- Makes concise statements with little extraneous information
- Establishes and uses conventional or standard speech (e.g., acronyms/shortcuts)
- Provides unsolicited responses (gives more detail than was asked, when appropriate)

Teamwork Dimension Rating Scales

ASSERTIVENESS Assertiveness refers to the willingness to make decisions, demonstrating initiative, and maintaining one's position until convinced otherwise by the facts. Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of assertiveness:		
Complete skill in assertiveness	5	Team members always bring up ambiguities or conflicts and ask questions when uncertain. Team members maintain their position when challenged, until adequate information is provided to persuade them. Team members make suggestions and often their opinion on decisions procedures, and strategies.
Very much skill in assertiveness	4	
Adequate skill in assertiveness	3	Team members sometime bring up ambiguities or conflicts and ask questions when uncertain. Team members sometime maintain their position when challenged. Team members make some suggestions and occasionally offer their opinions.
Some skill in assertiveness	2	
Hardly any skill in assertiveness	1	Team members never bring up ambiguities or conflicts and fail to ask questions when uncertain. Team members rarely maintain or defend their position when challenged. Team members do not make suggestions or offer their opinions.

DECISION-MAKING

Includes identifying possible solutions to problems, evaluating the consequences of each alternative, selecting the best alternative, and gathering information needed prior to arriving at a decision.

Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of decision-making:

Complete skill in decision-making	5	Team members gather information to evaluate possible solutions. Team members always evaluate each solution and explore consequences of the solutions. Team members cross-check information sources. Based on the information and consequences, team members select the best alternative. Team members periodically plan their activities and always follow the plans and decisions made.
Very much skill in decision-making	4	
Adequate skill in decision-making	3	Team members sometimes gather information to evaluate possible solutions. Team members sometimes explore the consequences of the solution. Team members rarely cross-check information sources. Based on the information and consequences considered, team members usually select the best alternative. Team members sometimes plan activities and usually follow the plans and decisions made.
Some skill in decision-making	2	
Hardly any skill in decision-making	1	Team members rarely gather information and explore the consequences of the solution. Decisions are made arbitrarily with little consideration of the information available. Team members rarely plan activities and may not follow the plans and decisions that are made.

SITUATION ASSESSMENT Situation assessment refers to the verbalization of information related to the perception of elements in the environment, the comprehension of their meaning in terms of task goals, and the projection of their status in the near future. Situation assessment verbalizations may serve to promote shared situation awareness between team members. Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of situation assessment:		
Complete skill in situation assessment	5	Team members frequently communicate situation assessment information. Team members immediately identify problem situations and recognize the need for actions. Team members notice differences in situation awareness from the other member and correct each other if needed. Team members actions and communications help to further an awareness of the ongoing mission status and ultimate goal.
Very much skill in situation assessment	4	
Adequate skill in situation assessment	3	Team members sometimes communicate situation assessment information. Team members sometimes identify problem situations and recognize the need for actions. Team members sometimes notice differences in situation awareness from the other team member and sometimes correct each other. Team members actions and communications usually help to further an awareness of the ongoing mission status and ultimate goal.
Some skill in situation assessment	2	
Hardly any skill in situation assessment	1	Team members rarely communicate situation assessment information. Team members usually miss problem situations and fail to recognize the need for actions. Team members rarely notice differences in situation assessment from the other team member and rarely correct each other. Team members actions and communications do not further an awareness of the ongoing mission status and ultimate goal.

LEADERSHIP

Team communication involves providing direction, structure, and support for other team members. It does not necessarily refer to a single individual with formal authority over others. Team communication can be shown by several team members.

Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of leadership:

Complete skill in leadership	5	Team members make statements regarding direction and strategy for the team. Team members listen to concerns of the other team member. Team members provide statements of team direction, strategy, and priorities. Team members set goals and follow through to meet those goals. Team members provide feedback to each other regarding performance.
Very much skill in leadership	4	
Adequate skill in leadership	3	Team members sometimes make statements regarding direction and strategy for the team. Team members sometimes listen to the concerns of the other team member. Team members sometimes provide statements of team direction, strategy, and priorities. Team members sometimes set goals and sometimes follow through to meet those goals. Team members sometimes provide feedback to each other regarding performance.
Some skill in leadership	2	
Hardly any skill in leadership	1	Team members rarely make statements regarding direction and strategy for the team. Team members rarely listen to each other's concerns. Team members rarely make statements of direction, strategy, or priorities. Team members do not set goals, or if they do, do not follow through to meet the goals. Team members rarely provide feedback to each other regarding performance.

COMMUNICATION

Involves the exchange of information between two or more team members in the prescribed manner and by using proper terminology. One purpose of communication is to clarify or acknowledge the receipt of information.

Rate the two-member team by circling the number which most closely represents the skill presented by the team in the dimension of communication:

Complete skill in communication	5	Team members acknowledge and repeat/verify critical information or information they did not understand. Team members always use accurate terminology. Team members establish and consistently use standard speech. Team members avoid extraneous communication unrelated to the task. Team members answer questions that are asked and provide further detail when needed.
Very much skill in communication	4	
Adequate skill in communication	3	Team members sometimes acknowledge and repeat/verify critical information or information they did not understand. Team members establish some standard speech and usually follow the standards. Team members rarely communicate on issues unrelated to the task. Team members usually answer questions that are asked and sometimes provide further detail.
Some skill in communication	2	
Hardly any skill in communication	1	Team members rarely acknowledge and repeat/verify critical information or information they did not understand. Team members don't establish standard speech, or, if they do, they fail to follow the standards. Team members sometimes communicate on issues un related to the task. Team members rarely answer questions when asked and do not provide further information than what was asked.

Subjective Rating of Perceived Workload
Teamwork and Taskwork Definitions

At the end of each trial, you will be asked to rate your workload during the trial. We would like you to try and rate the workload associated with *taskwork* separately from the workload associated with *teamwork.* For this purpose, please use the following definitions of *taskwork* and *teamwork.*

Taskwork

Taskwork refers to the workload associated with the individual tasks each team member must perform. Examples of taskwork within the TDT include: (1) monitoring the displays for information, (2) using the mouse and keyboard to work with the system, (3) analyzing the information displayed to make decisions required specifically for your role in the mission, etc.

Teamwork

Teamwork refers to the interactions between team members that are necessary for exchanging information, developing and maintaining communication patterns, coordinating actions, and negotiating decisions. Examples of teamwork within the TDT include: (1) communicating verbally with your team member, (2) communicating via the user interface with your team member, (3) coordinating the timing of actions with your team member, (4) working with your team member to develop communication standards or task strategies, etc.

NASA-TLX Workload Factor 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, controlling, 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 at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

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?

Frustration

How insecure, discouraged, irritated, and annoyed versus secure, gratified, content and complacent did you feel during the task?

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

APPENDIX B – SUBJECT SURVEY AND INFORMED CONSENT

Subject Survey

The intended purpose of this form is to establish a subject profile based on volunteered anthropometric data. Please complete the sheet to the best of your knowledge following the example formats indicated in the parentheses adjacent to each data field label.

Name:____

First   Last

Social Security Number (XXX-XX-XXXX): ______________________

Age (XX-yr.): __________ Gender (M/F): ______ Handedness (Left/Right): _________

Corrected Visual Acuity: Left Eye (XX/XX): ________ Right Eye (XX/XX): ________

	1	2	3	4	5
Video Game Experience:	None		Occasional		Frequent
PC Experience:	None		Occasional		Frequent

How well do you know your teammate?

1	2	3	4	5
Not at all	Acquaintance	Friend/colleague	Close friend/work together	Very close (spouse/roommate)

Do not write below this line. Experimenter use only.

Team # _____________   *Auto Condition: AS SHC DS BDM*

Role: AC IO

Informed Consent Form
Department of Industrial Engineering
North Carolina State University

THE EFFECT OF AUTOMATION ON TEAM PERFORMANCE AND TEAM COORDINATION

I hereby give my consent for voluntary participation in the research project titled, "The Effect of Automation on Team Performance and Team Coordination." I understand that the person responsible for this project is Melanie C. Wright, who can be telephoned at (919) 515-3295. She has explained to me the study objective of quantifying the performance and team coordination effects of different forms and levels of automation in a complex team task. Ms. Wright has agreed to answer any inquiries I may have concerning the procedures of the research and has informed me of my right to refuse to answer any specific questions asked of me. She has also informed me that I may contact the North Carolina State University (NCSU), Institutional Review Board for the Protection of Human Subjects by writing them in care of Dr. Matt Zingraff, Chair of IRB, Research Administration, NCSU, 1 Leazar Hall, Box 7514, Raleigh, NC 27695, or by calling (919) 515-2444.

Information concerning compensation for my participation in this study has been explained to me as follows: (1) I will receive $10.00 per hour for each of 2.5 hours resulting in a total compensation of $25.00 for my participation in the experiment. (2) In the event that I choose to terminate my participation in the experiment, I will be paid for only the time I have provided. (3) The researchers for the study have the right to terminate my participation if I am not cooperative or I experience discomfort or fatigue.

Ms. Wright has explained to me the procedures to be followed in this study and the potential risks and discomforts. In summary the procedures include: (1) an equipment familiarization period; (2) a dedicated training session(s) in the use of the Theatre Defense Task simulation; (3) a rest period; and (4) an experimental testing period. The experimental session will be scheduled within a single day. The session(s) will require approximately 2.5 hours of my time. The risks have also been explained to me as follows: (1) a potential exists for soreness of the hand and forearm muscles from extensive use of a mouse in controlling the Multitask© simulation and use of a keyboard in controlling the gauge-monitoring task; and (2) a potential exists for visual strain and/or fatigue in viewing the simulation displays through a conventional personal computer (PC) monitor. These risks are not substantially different from those associated with my everyday PC use. In the event that I experience fatigue or discomfort, I will inform the experimenters immediately.

I understand that if this research project results in any physical or mental harm to me, treatment is not necessarily available at the NCSU, Student Health Services, nor is there necessarily any insurance carried by the University or its personnel applicable to cover any such injury. Financial compensation for any such injury must be provided through my own insurance program. Further information about these matters may be obtained from the Institutional Review Board at (919) 515-2444, 1 Leazar Hall, NCSU Campus.

I understand that I will not derive any therapeutic treatment from participation in this study. I understand that I may discontinue participation in this study at any time without prejudice. I understand that all data will be kept confidential and that my name will not be used in any reports, written or unwritten.

Signature of Subject: **Date:**

Signature of Authorized Representative: **Signature of Witness to Oral Presentation:**

APPENDIX C – SUBJECT INSTRUCTIONS

Automation and Teams Procedures

Introduction

Thank you for volunteering to participate in this experiment. The goal of this experiment is to assess the effect of different forms of automation on the performance and coordination of teams. For this purpose, you will be asked to work as a team member on a Theatre Defense Task. This task requires you to make decisions that are similar to individuals in a military operation. You will be assigned to one of two roles for this task and will be expected to work with your team member to complete the task as best you can. The system you use may have some form of automation incorporated into it that will assist you with the task.

Overview of Procedures

The procedures we will follow today include:

(1) Completion of background and consent forms
(2) Training in the Theatre Defense Task
(3) Introduction to rating forms
(4) 15-minute practice session using the Theatre Defense Task
(5) NASA-TLX paired comparisons
(6) Four 15-minute experimental trials, followed by NASA-TLX ratings

You will be given breaks between each training session and each experimental trial. The entire experiment is expected to last between 2 ½ and 3 hours.

Completion of background and consent forms

Please complete the following informed consent form. This form summarizes the information that has been presented to you thus far and identifies the persons responsible for the study. The form also addresses University liability to the experiment. I encourage you to read the form. This form will not be associated with any of the other survey forms used in this experiment. In order to participate in this study you must have 20/20 or corrected vision and normal hearing. Please sign and date this form.

Please complete this background questionnaire. This form asks about your personal characteristics and will serve to verify your qualifications for the study. Please take a few moments to complete the survey. If you have any questions, I will be happy to address them. This form, like the informed consent form, will not be associated with any of the other survey forms used in this experiment.

[If applicable]. This is the payment form that will be used to calculate your compensation for participating in this experiment. Please fill out the information. Your Social Security number must be included on the form for tax purposes. The income you earn from this experiment is taxable and you should report it to the IRS.

Training in the Theatre Defense Task

During this task you will be asked to make decisions that are similar to ones faced by individuals during a military operation. You will be assigned to either the position of Air Commander (AC) or Intelligence Officer (IO). It is the Air Commander's job to destroy enemy aircraft. The AC will be viewing these aircraft on a display similar to a radar screen. It is the IO's job to determine which targets are enemies and provide this information to the AC. Your success will depend upon how well the overall team performs.

Training

You have been sent for training in your new job. During this training session you will receive in-depth instructions about your role and that of your teammate so that you can fully understand the task and work well together.

Workstation

Your station contains a set of radio headphones for communication with your teammate, your display screen, a keyboard, and a mouse controller. To use the radio headphones, you must press the button hanging on the wire when you wish to speak. You will not be able to hear incoming messages when the button is pressed. Please put on your headphones, test that you are able to contact your teammate, and adjust the volume as needed. While it may be possible for you to hear one another without headphones if you speak loudly, we ask that you imagine that you are actually located at different air bases and use only the radio headphones for communication.

Air Commander Display

The AC display includes a radar screen, a request information field, and scoring feedback. Home base is at the center of the radar screen. Aircraft will appear at the perimeter of the radar and travel inward towards your home base.

IO Display

The IO display has three main areas. At the top are the targets that have been requested by the AC for classification. In the center is the listing of targets that are currently available in the area, sensor readings associated with the targets, and the interface for classifying the targets and sending the information to the Air Commander. The bottom section of the display contains feedback regarding the outcome of targets.

The Task

[Demonstrate on AC Display]

Aircraft enter the radar screen from the outer ring.

[AWACS/Sensor reporting]

At the time that the aircraft enter the radar-scope, AWACS (or reconnaissance airplanes) make a determination of the expected aircraft type and send this information to the Intel Officer. The AWACS are displayed as small blue circles labeled A, B, C. The reliability of the AWACs information is dependent upon their location; the closer the plane is to home base (center of the radar), the more reliable is the information. At 0 to 10 miles out (first three rings) the reliability is 90%. At 10 to 20 miles (next two rings) the reliability is 60%. At 20 to 30 miles (outermost two rings) the reliability is 30%. It does not matter where the targets are within each range (a closer sensor within the 30% range is no better than one that is farther out in the 30% range).

Note that the determination of aircraft type is based on the position of the AWACS aircraft at the time the target enters the display, AWACS data is not updated during the time that the target travels toward home base.

[Demonstrate on IO display]

The IO considers the information provided by the AWACS sensors and makes a decision regarding the aircraft type. To do this, click on the 'View Information' button that corresponds to the target number or by pressing the number on the keyboard that corresponds to the button number on the screen. You may only see sensor data for one target at a time.

Once you have viewed the sensor data (and considered the reliabilities of the sources), you identify the aircraft type by selecting one of the aircraft type buttons on the right side of your screen.

[Automation, if applicable]

In addition, your station is equipped with automation that considers the position and target type to determine a priority order for classifying targets. Therefore, your list of targets is presented in priority order, with the most critical targets (targets believed to be enemies that are closest to home base) listed first. You may choose to use this information in determining what order to process targets in addition to considering Air Commander requests. There may be inefficiencies inherent in the automation such that the information presented may not always represent the best strategy, depending on conditions. Because of this automation, your list of targets continually re-sorts whenever a new target enters the display.

[BDM condition]

The IO display also contains a column labeled "Auto" that indicates what the automation determines the target to be based on the reliability of the sensor information.

[Demonstrate a classification]

When the IO presses a button to indicate a target classification, the target on the AC's display will change color. Friendly classifications appear as blue, green, or turquoise while enemy classifications will appear as red, yellow, or orange, depending on the target type.

[Show table]

It is the AC's job to let the friendly aircraft land at the base, and to destroy enemy aircraft before they reach the base. An aircraft will reach the home base if left alone (it will land on its own), unless it collides with another aircraft.

[Launching missiles]

You destroy aircraft by launching missiles at them. There are two types of missiles available; a Sparrow, which is smaller and faster and used for destroying fighters, and AMRAAM (armed medium range air to air missiles) which are long range radar guided missiles used for destroying bombers and transports. The left mouse button launches Sparrows (kills fighters or red targets) and the right mouse button launches AMRAAMs (kills transports and bombers or orange and yellow targets). The target will turn a light violet color to indicate that it has been targeted. If you launch an inappropriate missile at an aircraft the plane may not be destroyed and you will have to try again.

[Scoring, reference table again]

Point assignments have been made for each type of aircraft representing the target points for destroying the aircraft and the damage points for allowing the aircraft to land at your home base. The points are based on the mission relevance and lethality of the aircraft. For example, transports can carry many personnel and therefore could present a great threat if they land at a U.S. base, while the loss of a U.S. transport would be a devastating blow to our forces. Friendly aircraft have no score associated with getting through to home base and negative target points associated with destroying them. Enemy aircraft have positive points for being destroyed (target points) and negative points for reaching home base (damage points). Your goal is to get the highest number of target points by destroying as many enemy aircraft as possible (without destroying friendlies) and the lowest amount of damage associated with allowing enemy aircraft to reach home base.

[Collisions]

In addition, it is possible for some targets to collide with one another. If two friendly aircraft collide, no points are accrued for this. If two enemy aircraft collide the highest target points of the two aircraft are given. However, if a friendly and an enemy aircraft collide you receive double the damage points of the enemy aircraft. Since you get more points if you destroy two enemy aircraft on a collision course separately (as opposed to letting them collide) and the risk of damage points is great if an enemy and friendly aircraft collide, it is in your best interest to identify targets on collision paths quickly and destroy them if they are enemy aircraft.

[IO receives feedback on target]

The bottom part of the screen provides you with information regarding the outcome of the battle. It displays the target number, your classification, the actual target type and the final result for the plane ("destroyed", "got through", or "collided"). Incorrect classifications appear in red text. Correct classifications appear in blue text. Feedback for planes not classified appears in green text. The color-coding only indicates whether your classification is correct, it is not related to whether or not the outcome was correct (based on the AC's actions) for that particular aircraft. There will be a delay between making your classification and receiving this feedback, as the actual target type and outcome cannot be determined until the plane has either landed or been destroyed.

Communication

[Communicating reliability information]

[All conditions except Shared control]

The IO has no information regarding the reliability of the AWACS sensors, therefore, the AC must provide this information to the IO verbally. For example, the AC may report "A at 60%, B and C at 30%" or "A is highest reliability, use A".

[Shared Control]

The IO also has a listing of the reliability of the AWACS sensors [show on display]; however the IO will be very busy deciding on target classification and may not be able to keep track of the AWACs. The AC may choose to assist the IO by providing sensor reliability verbally. For example, the AC may report "A at 60%, B and C at 30%" or "A is highest reliability, use A".

[Communicating target information]

[Action Support]

The IO also has no information regarding the proximity of the target to home base or targets that are on collision course; therefore, the IO has no idea regarding what target is highest priority for classification. Consequently, the AC may choose to request specific target numbers.

[Shared Control]

The IO has a listing of target proximity to home base that may assist him or her in determining the order in which to classify targets [show listing]. This listing does not contain the same level of detail regarding distance from the center that is available to the AC. In addition, this listing does not provide any information regarding imminent collisions. Therefore, the AC may choose to request specific target numbers.

[Decision Support and BDM]

Since the IO's display is sorted with the nearest enemy targets at the top of the list, the IO has some information regarding the priority of targets for classification. This listing does not contain the same level of detail regarding distance of the targets from the center of the display that is available to the AC. In addition, this listing does not provide any information regarding imminent collisions. Therefore, the AC may choose to request specific target numbers.

Note, however, that because the automation sorts enemy targets toward the top of the IO's list, in some cases, the IO may have better information than the AC (who can not know the likelihood that a target is an enemy or friendly) regarding the priority of targets to classify.

[Target requests can be made through the interface (show AC interface) or verbally]

Target requests from the AC may be made verbally via the radio (e.g., "What is 56?").

Target requests may also be made by the AC through the user interface. To do this you simply enter the target number using the number pad or the numbers at the top of the keyboard. You will see them appear in the Request Information field. When the number is entered press the enter key and the information will be displayed to the IO. Note that if you enter a number that is not currently a target on the screen, it will not be sent and you must clear the number using the backspace key before you can enter another number.

[Communicating sensor data classifications]

When the IO decides on a target classification, this information can be transmitted through the interface (by pressing the appropriate button as shown previously) or it may be communicated verbally. For example “56 is an enemy.”

[Give list of standard terminology.]

A list of standard terms has been developed for use in this task. You are encouraged to use these terms, but are not required to.

Hints

[Work quickly the task is fast paced.]

Sensor data for aircraft will continually appear on your information screen until the test time is up. You could potentially process hundreds of planes in the time allotted for the task. Please be aware that this task moves quickly.

Also, because the target information displayed to the IO re-sorts frequently, if you do not move quickly in selecting a target, the currently active target may become in active.

[Consider the value of information in terms of point scoring.]

As far as accumulating points, the most valuable information the IO can provide is to identify targets as enemies, second to this is providing the actual target type, and identifying friendlies. If necessary, sacrifice the accuracy of the target type (as long as you have correctly identified friendly or enemy) rather than missing a target completely.

[Find ways to help each other, the IO is working harder (usually).]

Of the two roles, the Intel Officer is more demanding than the Air Commander. If you can find a way for the Air Commander to assist you (for example through verbal reports), then you should do so.

You will now have 5 minutes to practice in your role.

[Five minutes of practice time]

Now, please leave your headphones on your head, carry your radio in your hand, and change seats. You will now have 5 minutes to practice in the opposite role.

[Five minutes of practice time]

Please return to your original seat.

Rating forms

In order to assess the task workload that you experience during experimental testing, you will complete a subjective rating of various mental and physical demand factors. Here is a sheet of descriptions of each of the demand factors. Here is a sample of a rating form you will be asked to complete [show rating form]. In addition you will be asked to rate separately the workload associated with *teamwork* and the workload associated with *taskwork.* Here are definitions of teamwork and taskwork for you to use when making your ratings. The workload rating will be conducted in two steps. After you have completed a practice session on the Theatre Defense task, we will ask you to assess the importance of the various factors on workload for both teamwork and taskwork by comparing the different demand factors using this form [show paired comparison form]. Following the experimental trials, you will complete the ratings of each of the demand factors.

Following the experimental trials, we will complete rating forms that assess your performance as a team from the videotapes. We will be noting behaviors that represent good or bad teamwork. Based on these observations, we will rate you as a team on various characteristics that are associated with good teamwork. Note that we are not interested in identifying personal characteristics about you, rather we are interested in the effects that different levels of automation have on teamwork.

[10 minute break]

Practice Session

We will now begin a 15-minute practice session. Please use this practice session to familiarize yourself with the task and working with your team member. You are permitted to ask the experimenter questions during the practice session.

Now that you have completed the final practice do you have any remaining questions? You will not be allowed to ask any questions during the experimental trials.

NASA-TLX Paired Comparison

Please complete the workload factor comparison forms. First, indicate which of the two factors you believe is more important in assessing taskwork for the Theatre Defense Task. Now indicate which of the two factors you believe is more important in assessing teamwork for the Theatre Defense Task

[10 minute break]

Experimental Trials

Scenario

A U.S. base in the Middle East is currently under attack. The president has ordered all military personnel in the surrounding areas to be on alert for possible attacks. The IO is stationed at an air base in Western Europe. The AC is stationed at an air base in N. Africa. A few hours after the initial attack, both officers are instructed to report to their posts as enemy aircraft are making their way towards a U.S. base in N. Africa. Aircraft have been sent to intercept and destroy these incoming planes. It is your job to protect your base from enemy aircraft strikes while allowing your planes to land safely.

[Repeat for each of the four trials:]

You will have five minutes before taking over the command at your workstation to discuss strategy with your partner over the radio. Begin now.

On my signal, the Air Commander should respond OK to the prompt on the screen and the test trial will begin.

NASA-TLX Rating form

[Repeat for each of the four trials:]

Please complete the workload rating forms. First, indicate the level of workload associated with taskwork for each of the workload factors. Now, indicate the level of workload associated with teamwork for each of the workload factors.

[5 minute break between trials 1&2 and 3&4, 10 minute break between trials 2 and 3].

Debrief

Thank you for participating in this study. Do you have any further questions about the study? [If applicable], to obtain payment please take this form to Riddick 328-C and you will receive cash payment.

Target Types and Associated Scores

Plane Categories	Types	Color	Target Points	Damage Points
Friendly Fighters	F/A-18	Blue	-20	0
	F-15E	Blue	-40	0
	F-16	Blue	-60	0
Friendly Bombers	B-52	Green	-50	0
	B-1	Green	-80	0
	B-2	Green	-100	0
Friendly Transports	C-130J	Turquoise	-120	0
	C-21	Turquoise	-140	0
	KC-135	Turquoise	-150	0
Enemy Fighters	Mig-29	Red	60	10
	Su-35	Red	80	20
	Su-37	Red	100	10
Enemy Bombers	Tu-22M	Orange	10	50
	Tu-168	Orange	20	60
Enemy Transports	An-124	Yellow	50	50
	An-225	Yellow	60	60

Standard Terminology

It is common in jobs such as these that team members have been trained to used standard terminology to allow for more efficient and error-free communication. The following list of terms and their meanings have been given to both you and your teammate. You are encouraged to use these standard terms in your communications.

Term	Meaning
AC	Air Commander
IO	Intel Officer
Target	Unidentified Aircraft
Enemy	Target identified as an enemy
Friendly	Target identified as a friendly
Say again	Repeat the last message
Copy	I understood your last message
Reliability?	[Used by the IO] Provide an update of the reliability of the three sensors (A, B, and C)
Numbers (e.g., “Fifty-six” or “five-six”)	All numbers spoken without reference to something else are assumed to be target numbers.
“A”, “B”, or “C”	Refer to AWACS aircraft or sensors A, B, and C

APPENDIX D – EXPERIMENT DATA

Select this link to access a spreadsheet containing the performance, workload, communication count, and coordination rating data collected as part of the experiment.