2009-11-30

Using Operator Teams for Supervisory Control

Jonathan M. Whetten
Brigham Young University - Provo

Using Operator Teams for Supervisory Control

Jonathan Myrup Whetten

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Michael A. Goodrich, Chair
Charles D. Knutson
Robert P. Burton

Department of Computer Science

Brigham Young University

April 2010

ABSTRACT

Using Operator Teams for Supervisory Control

Jonathan Myrup Whetten

Department of Computer Science

Master of Science

Robots and other automated systems have potential use in many different fields. As the scope of robot applications that robots are used for increases, there is a growing desire to have human operators manage multiple robots. Typical methods of enabling operators to multi-task in this way involve some combination of user interfaces that support human cognition and advanced robot autonomy. Our research explores a complementary method of managing multiple robots by utilizing operator teams. The evidence suggests that for appropriate task scenarios, two cooperating operators can be more than twice as effective as one operator working alone.

ACKNOWLEDGMENTS

I am grateful for the wonderful guidance and encouragement from my advisor, Dr. Michael A. Goodrich. I could not have completed this research without his support and mentoring. I appreciate the environment that exists in the BYU Computer Science Department that allowed me to expand my love of learning. I would also like to thank my parents, Delon and Loraine Whetten, for being supportive through all of the educational endeavors in my life. Most of all, I am very thankful for my beautiful wife Jennifer, who stuck with me and gave encouragement even when I was making no money, never spent much time at home, and would not stop talking about my research when we were together.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computers are revolutionizing the way we deal with information, and mobile robots hold the promise of extending this change to our physical world. In this thesis we discuss current approaches to human-robot interaction (HRI) that allow operators to monitor and control multiple remote robots. We then explore a method of improving operator performance on a search and navigation task through the use of operator teams. The benefits of operator teams are applicable to the general HRI problem of increasing the number of robots that human operators are able to simultaneously monitor and control.

1.1 Human-in-the-loop with a large number of remote robots

Due to robot autonomy, human operators are able to change roles from robot micro-management to more of a supervisory role (for a discussion of supervisory control, see [73]). This switch enables operators to control multiple robots, since their attention is not consumed by managing just one robot. A significant amount of research into HRI has been to continue to improve supervisory control in order to help one operator control as many remote robots as possible.

Automating previously manual tasks can change the task from active operation and control to passive monitoring, which can cause problems. Operators may experience skill degradation or unevenly distributed workload [3]. Operators have a common problem of calibrating a proper amount of trust in the automation; too much trust in the automation causes preventable errors, while too little trust results in reduced efficiency and possible operator errors [50]. Automated systems can also "camouflage" a dynamically evolving problem by compensating for it silently until it is too big of a problem for either the automation or the operator to recover from [3]. Understanding the capabilities and limitations of various automated modes is a significant issue for operators, as well as knowing what mode the autonomy is in [71].

Chen et al. [14] point out that a manual mode in which robots take no initiative of their own and follow only the commands they are given will be a requirement for remote robots in the foreseeable future. For all but the most mundane tasks, even so-called fully autonomous robot will require a human-in-the-loop setup, even if only to provide high-level goals and decisions or to reprogram and maintain. Because there is a human "in the loop," human shortcomings and frailties must be addressed to ensure effective human-robot systems. The number of tasks that operators can perform simultaneously is restricted by cognitive limitations. Finding ways to assist human cognition in managing robotic systems is an important area of research in the HRI field.

Prior work has focused on interface design and collaboration methods between operators and robots to improve supervisory control [60, 61, 7, 65, 35, 14]. Instead, we explore a method of reducing cognitive workload that is overlooked in existing literature: operator teams.

1.2 Perspectives on Operator Teams in Other Problem Domains

Working in teams has significant benefits to many different areas. Throughout human history whenever a problem is beyond the ability of one person, humans tend to find a way to approach the problem as a group. We organize ourselves into groups and divide up responsibilities to manage the complexity of the world around us. Business organizations use teams to brainstorm and evaluate decisions. Software project managers use teams to break up projects into small enough pieces for individuals to handle.

Using teams to operate remote robots in real-time is slightly different from these situations because it imposes strict temporal constraints on decision-making and planning. A better example of teams under temporal constraints and high workload would be airplane crews and air traffic control [79]. Air traffic controllers commonly work in teams with one controller dealing with communications and radar monitoring tasks, and the other dealing with flight data. Air traffic controllers can also be viewed as supervising many highly automated robots, and air crews are called on to jointly monitor the large number of automated systems that now exist on commercial airplanes.

Teamwork has been recognized as necessary in airplane cockpits to the extent that formal training procedures have been established for airplane crew interpersonal communication [47]. The task of flying a commercial airplane is too much for one person in an emergency situation, and cannot be broken down into tasks that can be performed independently. Given the high cost of error for flying a passenger plane, having at least two people who can fly the plane is also the norm for the sake of redundancy (the recent death of the captain of a Continental Airlines trans-Atlantic flight is a good example of the need for this redundancy [4]).

In contrast, robots are considered expendable in comparison to human life, and so some error may be tolerable from operators, reducing the need for redundancy. The task of controlling multiple robots can usually be broken down into subtasks (controlling an individual robot, for example) that can be performed independently of one each other. Because of these differences, it is not readily apparent that remote robot operation can benefit from teams in the same manner that piloting commercial aircraft does. This thesis shows that operators working in team can be more effective than operators working alone.

1.3 Thesis Statement

Given the proper task domain, interface support, and robot autonomy, a team of two operators can be more than twice as effective as a single operator when controlling remote robots in a navigation and exploration task. In the absence of these conditions, two operators can be less than twice as effective as a single operator.

1.4 Thesis Organization

This thesis is organized as follows. Chapter 2 discusses relevant existing research relating to teams, cognitive processes, and robot control and management. Chapter 3 presents the experiment that was performed to validate our hypothesis. Chapter 4 analyzes the experiment and identifies the improvements gained by operators who worked in teams compared to those who worked alone. Chapter 5 summarizes findings and examines questions raised by this work and possible future applications of it.

Chapter 2

Related Literature

In this chapter we review relevant research in areas relating to robot operator performance and team dynamics. These topics serve as a backdrop for how operators working in teams are able to be more effective than operators working alone, and what can be done to exploit the advantages operator teams provide.

2.1 Cognitive Processes

The effectiveness of remote robots is largely dependent on the mental capabilities of human operators. In order to understand how to improve human operation of remote robots, we must first understand the limitations imposed by human cognition.

2.1.1 Workload

Intuitively, we know that humans have limitations on the amount of information we are able to process and the number of tasks we can perform simultaneously. These limitations are generally framed formally using the related concepts of mental workload, working memory, and attention resources.

In general, remote robot operation suffers from high workload, though as supervisory control becomes more widespread, low workload may pose just as significant a problem [3]. In controlling multiple robots, workload is a significant concern because the wrong level of workload will reduce the benefit of having more robots available.

There is little agreement on a formal definition of workload, but it can be intuitively understood as the mental work someone puts into a given task. For example, solving complex math equations requires more mental energy, and therefore causes higher workload, than holding a casual conversation.

The effects of workload on human performance vary greatly between tasks and people depending on the task at hand, the skill and experience of the individual, and external factors. In addition, humans have shown the ability to dynamically increase workload capacity by focusing their efforts when faced with increased task load [48, 42]. Because of these factors, efforts to empirically discover the limits of mental workload (the so-called workload "redline") have generally come up empty-handed, though for specific domains there has been some success (for example, see [15]).

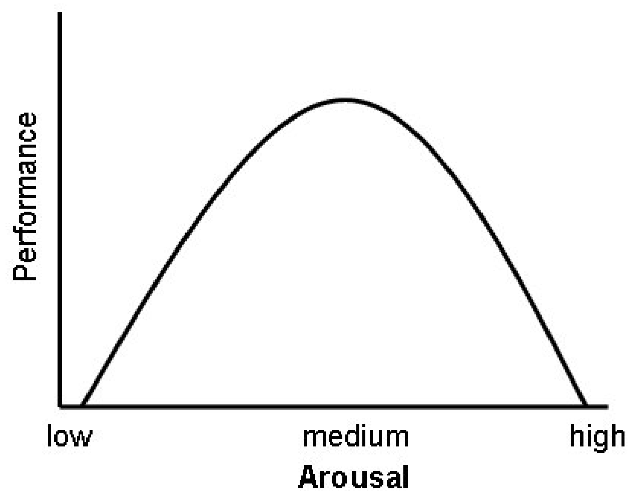

Figure 2.1: The Yerkes-Dodson curve.

Human performance declines when attention or arousal is too low, as in cases of excessive boredom, as well as when arousal is too high. These ideas are based on the Yerkes-Dodson curve (see Figure 2.1), so named for work done by Robert Yerkes and John Dodson in which they recorded the learning rate of mice in relation to varying amounts of stimuli. Subsequent studies have consistently shown similar relationships. Hebb [46] found that the Yerkes-Dodson curve relates to arousal and task performance, Cummings and Guerlain [19] found a similar link between the frequency of task events and performance, and Cummings and Nehme [20] demonstrated that the relationship exists between performance and supervisory control of multiple agents.

People generally use one of four strategies to adapt to high workload [78]. 1) People allow task performance to degrade. 2) People perform tasks more efficiently, either bringing to bear increased focus and attention, or changing from finding optimal solutions to satisfactory ones using heuristics. 3) People may shed tasks altogether, eliminating least important tasks first. 4) People shed tasks in a non-optimal fashion, eliminating tasks that should be performed (commonly a result of a lack of expertise).

Another strategy to deal with high workload is to have a team perform a task instead of individuals. In certain situations, people may extend themselves too far, choosing to allow their task performance to degrade rather than shedding tasks. In the case of operating multiple robots, this situation can be alleviated by participation in a team, where task demands can be shared among team members.

2.1.2 Situation Awareness

In order to operate remote robots effectively, the operator needs to know about the environment around the robot. People filter through an enormous amount of information based on what they perceive from their own environment, and use this information to plan, make decisions, and come up with contingencies. The planning and decision-making requirements for operating a remote robot are at least as complex as they are if the operator was physically performing the task themselves instead of the robot, but the process of gaining the information necessary is severely constrained. For example, Woods et al. liken operating a remote robot to driving while looking through a 'soda straw' because of the limited field of view available [81].

The result of perceiving and filtering information from the world around us (or in the case of remote robot operations, the world around the robot) is called situation awareness. Situation awareness is a concept that seems to defy a simple definition. Though the concept of having a knowledge of the environment you are working in has an immediate, intuitive meaning [70], a clear and concise definition has yet to gain widespread acceptance. Probably the most popular definition is the one given by Micah Endsley: "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" [27].

While this is an effective definition for most purposes, it leaves out the cyclical nature of situation awareness. Dominguez [23] combines the definitions from Endsley and from Carroll [10] to produce the following definition of situation awareness, which more strongly emphasizes its process aspect: "Continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events."

The difference in viewpoints seems to stem from describing situation awareness as a process as opposed to a state. Endsley states that "it is important to distinguish the term situation awareness, as a state of knowledge, from the processes used to achieve that state. These processes, which may vary widely among individuals and contexts, will be referred to as situation assessment or the process of achieving, acquiring, or maintaining SA" [26]. Flach [34] cautions that modeling situation awareness as an object similar to working memory can lead to circular reasoning where situation awareness is the cause of itself, which is not helpful in any practical way.

According to Endsley, situation awareness has three levels: 1) perception of the environment, 2) comprehension of the information available to "derive operationally relevant meaning" from available information, and 3) projection of the current state of the environment into the future in order to anticipate future events [29]. She further models situation awareness as having cognitive processes, effects that can interfere or support those cognitive processes, and a memory model where our perception of the environment is stored and accessed [29]. The relevant cognitive processes include perception, attention and pattern matching with previous similar situations, each of which has been studied extensively in their own right. Factors that can interfere with (or promote) these processes include stress, workload, interfaces, task complexity, and automation. In Endsley's model, our interpretation of the environment is then colored by our goals, by our expectations, and by the mental models we have built up in past situations.

All of these factors contribute to the final product, which has been described variously as a "coherent internal world view" [40] and "a holistic picture of the environment" [25]. This "world view" is then used in decision making and planning, allowing a person to take their surroundings into account when taking action.

Situation assessment, or the process of gaining situation awareness, can be very time consuming, particularly in challenging environments. While individual cognitive processes are often measured in the sub-second range (for example, see [53]), complete situation assessment can take much, much longer. In a simulated search and rescue task using a remotely operated robot, Burke et al. [9] found that 54% of operator statements (which included statements to team members) were related to gaining or maintaining situation awareness and that the robot was stationary for nearly 50% of the time (compared to actively searching, presumably because operators were consumed with establishing situation awareness). Similarly, Yanco and Drury [83] found that operators spent 30% of their time solely in activities related to gaining situation awareness in an urban search and rescue scenario.

Predictably, it seems that greater situation awareness can lead to better performance, especially if the process of situation assessment is sped up. Chadwick et al. [13] found that robot operators completed a navigation task significantly faster when presented with a third person view of the robot and its environment (which allows for more of the scene surrounding the robot to be viewed and increases situation awareness) when compared to using a first person view. Murphy [57] goes even further and claims that better mobility and navigation in mobile robots will reduce the time spent on a mission by no more than 25%, unless they are also accompanied by an improvement in situation awareness support.

When controlling multiple robots, the impact of situation assessment increases. Since situation awareness must be maintained for multiple robots in many different locations, operators are unable to bring their full mental capabilities to bear on a single robot without completely ignoring the rest. An overly simplistic analogy would be trying to use multiple programs on a computer which is only capable of holding one of the programs at a time in memory. For modern computers, this state is called *thrashing* because more time is spent loading information into memory than is spent doing actual work. Similarly, it is easy to picture an operator spending far more time gaining situation awareness for multiple robots than they spend giving the robots commands (in fact, this already happens with single robots).

Operator teams can lower the amount of time needed for situation assessment by allowing individuals to maintain more information about the state of the robots they are controlling within working memory. The more state information that is retained from previous interaction with a robot, the quicker the process of situation assessment.

2.2 Fan-out

Human involvement in robot operation will exist for the forseeable future, but advances in automation will allow operators to monitor many robots at once. The number of robots that can be controlled is modeled using fan-out [38]. Fan-out is a metric that represents the number of simple homogeneous robots a single operator can manage. It is calculated based on how long the automation can go without intervention (neglect time) and how long it takes an operator to input new commands or correct problems with the system (interaction time). The formula for fan-out is:

$$Fan\ Out = Neglect\ Time\ /\ Interaction\ Time + 1$$

While it is fairly simplistic, this model has been empirically validated by Olsen et al. through a series of user studies [63]. This model of fan-out has significant limitations, including validity for only simple homogeneous robots and tasks. Further work by Cummings et al. looks at extending the original fan-out equation to include wait times and performance metrics to try to model the "optimal" level fan-out for a given set of task constraints [21]. Goodrich et al. look at including switching costs and extending the model to include heterogeneous robot teams [39]. Based on the original formulation, we can seek to increase fan-out by either increasing neglect time or by decreasing interaction time.

2.2.1 Neglect Time

Neglect time is a measure of how long a robot can be effective on its own without human intervention [38]. It roughly corresponds to the level of autonomy used by a robot, and so increasing neglect time can usually be achieved by improving the intelligence and capabilities of autonomy.

Many areas of research are being pursued that can increase neglect time, such as swarm robotics, supervisory control, mixed initiative robots, adaptive autonomy, adaptable autonomy, and more. Some of these approaches can be used in conjunction with one another in robotic systems, but none of them solves the problem entirely. Generally speaking, good autonomy is difficult to design, domain-specific, and often flawed in non-trivial ways.

The interaction between human operators and autonomy can be complex and unpredictable. Endsley [28] states that automation affects situation awareness in at least three different ways: 1) operator complacency levels due to a monitoring role, 2) changing an active task to a passive one, and 3) changes in the feedback given to the human operator. Wiener mentions that automation can often change the type or timing of errors instead of eliminating them, and that it can also cause an uneven distribution of workload [80].

Without significant neglect time, control of multiple robots is counterproductive because only robots which are actively receiving attention will be effective. Because of this, increasing neglect time through automation is an important goal, but only part of the solution.

2.2.2 Interaction Time

The other way to increase fan-out is by decreasing interaction time. Interaction time varies greatly between operators and is even more difficult to measure than neglect time (in fact, Olsen et. al. found it easier to measure fan-out and neglect time and then calculate interaction time [63]). Many factors contribute to interaction time, including time spent switching between tasks, time to gain situation awareness, time to plan, and time to communicate that plan to the robot [38]. Interaction time can be broken into two basic categories: the system time required to give commands through the interface and actively retrieve the information that is needed, and the cognitive time required to mentally process information and make decisions.

Reducing system time is a fairly well understood, if still difficult, task. Factors that affect system time are primarily executing commands and navigating through menus. Many well-known principles from user interface design apply here including Fitt's Law [33], the Steering Law [1], and others. System time can be minimized through good interface design.

The cognitive aspect of interaction time is much more difficult to quantify and improve, but there is evidence to suggest that an enormous amount of cognitive processing goes on in the process of gaining situation awareness [9, 84], and that the process can be assisted by proper interface design with an eye to improving situation awareness [13, 61]. Other important considerations include task switching [53, 12] and workload [20, 78].

Interaction time is most likely to be the bottleneck in robotic systems as autonomy continues to improve. Indeed, Murphy [57] says that the time needed to gain situation awareness is already a significant hindrance for search and rescue.

Operator teams allow individuals to focus on managing subsets of robots that are somehow related to each other, either spatially or logically. This relationship can potentially reduce the cost of interaction time by reducing switching time and speeding the process of situation assessment. However, working with a teammate requires coordination, which effectively introduces a new task which requires additional effort and attention from team members. Our work explores the questions of (a) whether the benefits of operating robots as a team exist, and (b) if these benefits outweigh the added effort required to coordinate within a team.

2.3 Teams

In order to create effective operator teams to control multiple robots, some fundamental knowledge is required about how teams operate, how they are organized, and what factors affect their performance.

Teams and teamwork have been studied extensively for several decades because of the impressive results that can be obtained by utilizing them effectively. Due to the sheer number of researchers looking at the problem over the years and the inherent complexity of trying to understand how human beings work and think, there are many different models of how teams initially form, how they operate, and what their benefits are. We explore a few of those issues here. Of particular interest are the types of teams that exist and the types of problems for which they can effectively be utilized.

2.3.1 Unitary and Divisible Tasks

In his seminal work, Steiner [75] explores team process and limiting factors on productivity. He separates teams into a taxonomy based on the type of problem they are attempting to solve, rather than team composition or organization. He primarily focuses on small "task groups" that exist to achieve a specific goal within an allotted amount of time.

Steiner broadly categorizes teams based on *unitary tasks* or *divisible tasks*. Unitary tasks are those that cannot be profitably divided into smaller subtasks. They will not receive measurable benefit, and may in fact be hampered, by having additional people join in completing the task. Unitary tasks are mainly classified as *disjunctive*, *conjunctive*, or *additive* [75]. Disjunctive tasks are ones in which the performance of the group is limited to the ability of the most capable individual for that task within the group. Performance on conjunctive tasks is limited to the ability of the *least* capable member for that task. Additive tasks provide group performance that is a cumulative total of the performance of each individual member. Unitary tasks in general are not necessarily well suited for groups, and provide diminishing returns compared to individual effort.

Divisible tasks are those in which a division of labor can have clear benefits, and therefore be profitably performed by a groups instead of individuals. A common example of this is assembly lines, where different portions of the work are done at separate stations and workers become specialists. Though there is a potential for significant payoffs for doing tasks as a group, there are also the added nuances of group interaction and other factors which can easily counteract any potential benefits. Steiner states that most real-world tasks are divisible, and classifies them based on how tasks are divided into subtasks and how people are assigned to these subtasks.

One of the shortcomings of Steiner's work is that he focuses mostly on teams dedicated to doing manufacturing and other types of physical labor. There is mention of certain group mental tasks, such as brainstorming or solving a logic puzzle together, but the research in groups prior to 1972 had been done primarily in military or industrial settings. Since remotely operating a robot is primarily a mental task, operator teams do not necessarily fit well into Steiner's model.

In contrast, Fisher and Fisher [32] discuss teams from a purely knowledge-work perspective. They provide four basic categories for knowledge-work teams: natural work teams which form based on the requirements of the job, cross-functional teams which have purposes that cross multiple natural work teams, small project teams that exist for a specific task and then are disbanded, and special purpose teams which are something of a combination of cross-functional teams and small project teams.

Fisher and Fisher describe the key differences between physical work and knowledge work (replicated in table 2.1). Looking at the differences, supervisory control falls largely into the knowledge work category, with the exception that the work outcome is a tangible product; the physical labor is done by the robot.

	Physical Work	**Knowledge Work**
Core task	Doing	Thinking
Critical skills	Physical	Mental
Work process	Linear	Nonlinear
Knowledge used	Applied	Created
Work outcome	Product	Information

Table 2.1: Adapted from [32].

2.3.2 Human-robot Teams

In spite of apparently fitting into the category of knowledge work, operating remote robots seems in many ways to be more similar to production work roles. In both remote robot operation and production work, there are specific goals to be accomplished and a relatively small time-frame in which they can be accomplished, and there is strong temporal pressure since activities are occurring without regard to the operator.

Within the HRI field there has been significant research into different types of human-robot team configurations and their benefits. A fairly comprehensive list of possible human-robot team setups was compiled by Yanco and Drury [82] and is replicated in Figure 2.2. The groupings in Figure 2.2 imply teams where decisions are explicitly coordinated among all team members. Human teams in this sense would agree upon a command or task to send to the robot(s) and robot teams would receive commands as a group and autonomously decompose the command into specific tasks for individual robots. The arrows between items in Figure 2.2 indicate a relationship in which some form of communication and/or coordination takes place.

Figure 2.2(A) specifies the classic one-to-one relationship between human and robot, which is still the most common configuration. The automation support for many common tasks like explosive ordinance disposal is not enough to allow operators to control two robots at the same time. In fact, often the task is complex enough or the risks and costs of failure high enough that human teams control one robot by agreeing together on commands before they are sent to the robot(s) as in 2.2(B). This is what is done currently with Mars rovers, and two-person teams controlling one robot seem to be far more effective than a single operator in search and rescue tasks [58, 9]. Multiple individuals controlling one robot (as in 2.2(C)) was demonstrated by Goldberg et al. by allowing distributed, collaborative control over a robot arm by anonymous individuals connected through the Internet [37].

Robot teams shown in Figures 2.2.(D)-(F) imply coordination on the part of the robots to execute commands given to the team as a whole. This is used in the field of swarm robotics

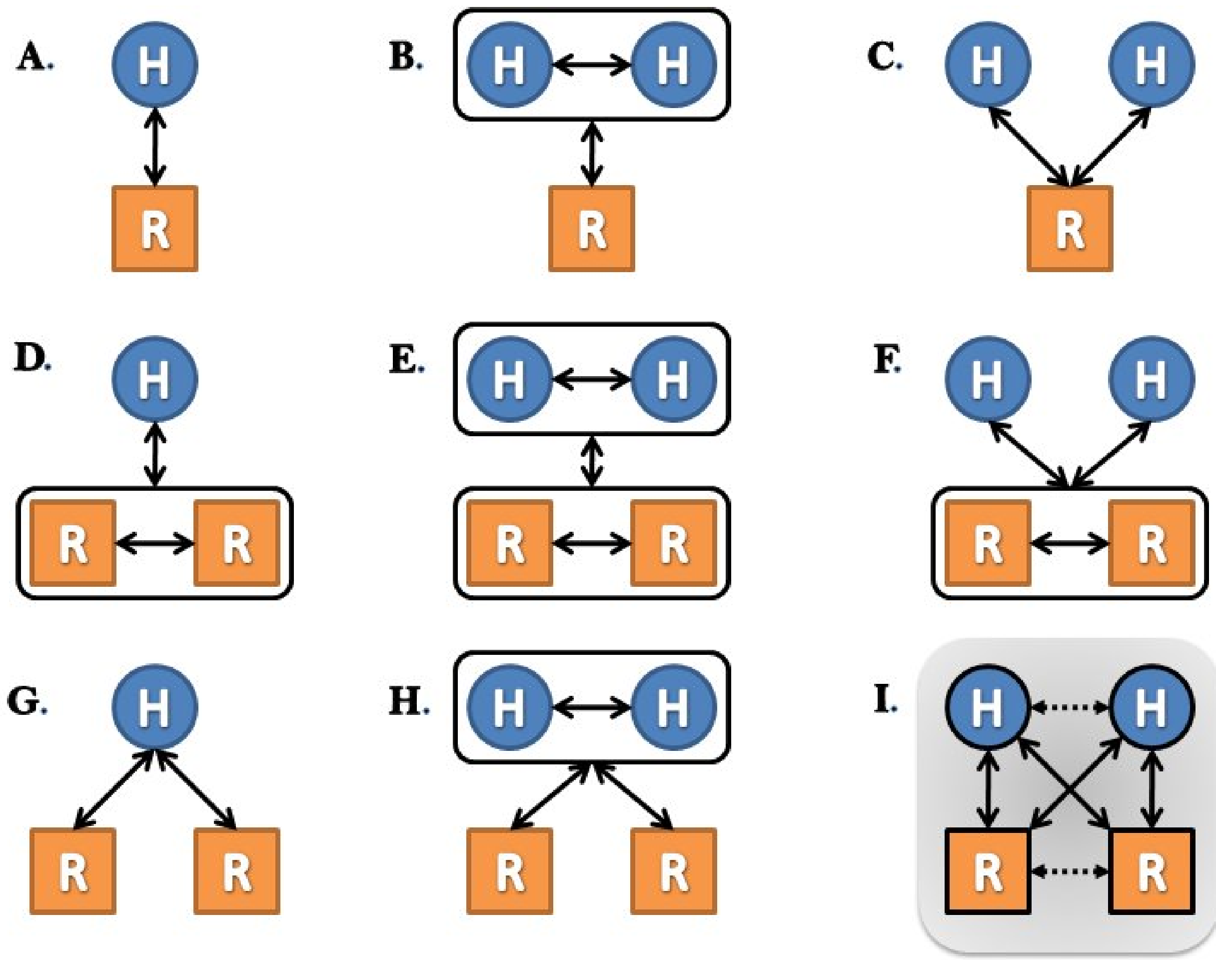

Figure 2.2: Shows various configurations for human-robot teams, with the boxes indicating coordinating teams. Adapted from Yanco and Drury [82], with the notable exception of Fig. I., which they intentionally excluded.

[5, 11, 51], but is also implemented at the interface level through the use of “playbook” commands that result in more specific behaviors being implemented in different robots [65].

The holy grail of supervisory control is to form human-robot teams modeled after Figure 2.2(G), where one operator controls a large number of remote robot. This has not yet happened primarily because sufficient levels of autonomy have not been developed to allow swarms to be effective with high-level commands, and sufficient support for reducing interaction time has not been developed to allow many different robots to be controlled by one person working alone.

Figure 2.2(I) is the model used for our experiment. There is no enforced coordination between robots or operators and they task the robots separately. The dashed lines represent the fact that communication occurs between operators, unlike the Internet connected robot in Goldberg et al. [37]. In a real system there would be communication and coordination between robots as well, even if they are individually tasked. This model was disregarded in Yanco and Drury's taxonomy as being unworkable. Nickerson and Skiena [59] describe a model similar to this that they liken to "call centers," suggesting that this could be a way that the robot-to-operator ratio can be increased. Lewis et al. took this suggestion further and simulated a three-tiered system in which a commander gives high-level goals to a group of operators (similar to the call-center approach) who then send commands to robots under their command [52]. Their experiment only simulates human decision-making, but their finding indicate that this is a workable implementation.

2.3.3 Benefits and Limitations of Multiple Robots

Dudek et al. [24] discuss models of robot teams, and the types of tasks robot teams might perform. They claim that there are tasks that use multiple agents, similar to the requirement of two people turning two keys at the same time some distance apart in order to launch a missile, and that there are other tasks that use a single agent, meaning that one robot performing one task in one location would not likely benefit from having additional robots to assist. In between these two extremes are tasks which *might* benefit from using multiple robots.

Dudek et al. state that the draw for multiple robots is two-fold. First, having a collection of robots working on the same task can improve reliability of the robotic system as a whole. If one robot is incapacitated, others can fill in for it. Second, the task may be performed more efficiently if there are multiple robots present. Dudek at al. state that at best using multiple homogeneous robots will improve efficiency in a linear fashion based on the number of robots, but that in practice this theoretical limit will not be reached due to efficiency lost by the need to coordinate effort.

While it may be true that linear growth in efficiency with increasing group membership is the most we can expect from robot teams (though this claim may be suspect), the same argument cannot be made for human teams. In their book, Fisher and Fisher [32] share many anecdotes of how teams improved the performance of various companies and organizations, and more applicable to our discussion, Murphy and Burke [58] showed a ninefold increase in effectiveness over one operator by using two operators to control one robot.

2.3.4 Productivity and Team Synergy

In order to exploit the benefits of working in teams in order to achieve greater than linear growth in performance, we first need to understand what performance gains are available to teams to begin with.

In 1972, Steiner [75] provided a formula to calculate the productivity of groups:

Actual Productivity = Potential Productivity - Losses Due to Faulty Process

where potential productivity is calculated from a combination of task demands and resources available. This formula represented significant progress for the study of teams and team process at the time. Steiner's model focuses on a maximum (potential) productivity which he says can theoretically be calculated before a group actually starts work. Any mismatch between actual performance and this maximum is caused by "losses due to faulty process," or the costs of working in a group.

Many costs have been associated with groups over the years, but several stand out as unique and common (summarized in Table 2.2). *Production blocking* occurs when the group organization requires that some groups members stop contributing. For example, when one person is speaking to the group, the rest of the group needs to listen and cannot (productively) talk simultaneously [62]. Group members often suffer from *evaluation apprehension* in which contributions are not made for fear of a negative reaction from other group members [69]. *Social loafing* or *social impairment* is a significant phenomenon where members of a group give as much as 50% less effort than they would working as individuals [66]. The effort seems to decrease as individual accountability decreases and as group size increases [49]. *Cognitive interference* occurs when the ideas or contributions from other group members interrupt cognitive processes for individuals, lessening their contribution [67]. *Communication speed* affects coordination time and can increase the effort needed to communicate. This increased communication effort can hamper team coordination and adversely impact performance [54].

Production Blocking	Organization overhead, or the need to "take turns" in a group.
Evaluation Apprehension	Individuals fear judgment from other group members and so withhold their own contribution, or are overly conservative.
Social Loafing	Also called the "free rider problem" [77]. People have a tendency to "hide" in groups and put forth less effort.
Cognitive Interference	The actions or statements of others cause an individual to forget or be less effective with their own contribution.
Communication Speed	How long does it take to coordinate? This is affected by the communication modality, whether it be speech or typing or something else entirely.

Table 2.2: Potential causes of loss in efficiency in groups.

While the insight into process loss is significant, the difficulty with Steiner's formulation is that he does not provide a way to model the *gains* from working in a group. This is complicated by the fact that process gains have been notoriously hard to reproduce and measure in laboratory settings, and so proof of their existence is sometimes no more than an intuitive understanding of how groups operate. At least two specific group process gains have been noted in the literature: the *social facilitation effect* and the *assembly bonus effect*.

The social facilitation effect is based on the observation that performing a task in the presence of others can cause performance to significantly increase [86]. This is not a universal effect, however, and it seems to directly contradict social loafing. It appears that the deciding factors of whether individual performance will increase or decrease within a group include group size, consequences of performance (greater reward corresponds with greater effort), task complexity, and individual performance identifiability [43, 66].

The assembly bonus effect occurs when group interaction combines members' knowledge in such a way that higher quality decisions are made by the group than could be done by the group's best member [16]. Though this has largely been accepted as conventional wisdom, the effect has been disputed many times due to the difficulty in satisfactorily recreating it in a laboratory setting (see for example Michaelson et al. [55], whose work was challenged by Tindale and Larson [76], whose assertions were in turn challenged by Michaelson et al. [56]). It seems that the exact nature of the assembly bonus effect is still an open question [68].

Such positive gains from using groups generally fall under the umbrella term of "synergy." Synergy can be loosely defined as the benefits received through cooperation that exceed the contributions of individuals. Though synergy has been equated with the assembly bonus effect [22], positive group effects like assembly bonus and social facilitation are a subset of synergistic gains available to groups. It is a mistake to use the terms interchangeably.

In the rush to realize the benefits of synergy, some organizations have misused teams [41], causing synergy to be disparaged as a "buzz word" and mocked in popular media such as Scott Adams' comic strip *Dilbert* [2]. Such sentiments notwithstanding, as Corning [17] states:

> "Synergy is real. Its effects are measurable or quantifiable: e.g., economies of scale, increased efficiencies, reduced costs, higher yields, lower mortality rates, a larger number of viable offspring, etc. More subtle measuring rods include enhanced stability properties, greater stress tolerance, increased fidelity in reproduction, the melding of functional complementarities to achieve new properties, and so forth."

Synergy is used in the medical field to describe positive drug interactions, such as combinations of antibiotics that perform better together than any individual drug on its own [6]. Synergy exists within colonies of bacteria that allow them to secrete enzymes in sufficient quantities to be able to digest food [17]. It is even applicable to evolutionary fitness and the development of altruism as described by John Maynard Smith's "Haystack" model [74].

We have discussed two forms of synergy in relation to group benefits, but there are many different types of synergy mentioned in the literature. Steiner [75] focuses on the division of labor and its associated benefits. Goold and Campbell [41] suggest that the benefits of synergy to business units include shared knowledge, coordinating strategies, shared resources, vertical integration, and economies of scale. Felin and Knudson argue that it is for these types of benefits that organizations exist at all [31].

Corning defines synergy slightly different than we have, saying that "The term is frequently associated with the slogan 'the whole is greater than the sum of its parts'... We prefer to say that the effects produced by wholes are different from what the parts can produce alone" [17]. Using this more liberal definition, Corning provides many different types of synergy. “Additive phenomena” has benefits that grow linearly with the number of parts. “Emergent phenomena” are exhibited when many parts combine to form something different from any individual component, as is seen quite often in metallic alloys like stainless steel. Division of labor is an important class as well, as there are many examples of specialization in both humans and animals. Corning goes on to find examples of synergy in fields ranging from quantum physics to neurobiology.

Our main interest in synergy is restricted to our operational definition in which systems can achieve greater than linear gains as the number of members is increased. These synergistic relationships exist and can be exploited. Our goal is to use synergy to increase the number of remote robots that can be controlled by a team of operators working together instead of as individuals.

Chapter 3

Methods

In order to show empirically that two operators working together can be more than twice as effective as one operator working alone, we conducted a 2x2 experiment using novice operators. The independent variables we used are the number of operators and the number of tasks operators were given to complete. We gathered data from human test subjects with eight test runs in each of the four conditions.

3.1 Independent Variables

For each of the two independent variables there were two conditions (see Table 3.1). Subjects were asked to do one or two primary tasks, and they either completed the tasks by themselves or with a teammate. Following our hypothesis, we expected to see significantly improved performance in tests with two tasks and two operators compared to tests with two tasks and one operator.

Experiment Conditions	
One Operator One Task	Two Operators One Task
One Operator Two (sub)Tasks	Two Operators Two (sub)Tasks

Table 3.1: Independent variables.

3.1.1 Number of Operators

Our experiment setup allowed a team of two operators to see the same virtual workspace, as well as the results of each other's commands. Beyond this, there was no coordination enforced, but communication was allowed through a chat window. Robots acted independently of each other, and the operators did not have to agree upon commands to go to the robot, and in fact could issue opposing commands which were resolved by having the robot simply following the last command received.

Using multiple operators is not a new idea. Murphy and Burke [58] found a ninefold improvement in effectiveness using two operators to control one robot. Yanco and Drury [82] discuss a range of human-robot team structures, many of which involve operator teams. The Mars rover is controlled by a team who carefully decides what commands to send to the robot. In spite of this, our team organization differs in a few significant ways.

In spite of being thorough in their discussion of possible human-robot teams, Yanco and Drury [82] intentionally left out the condition in which multiple operators controlled multiple robots without explicit coordination on either side, since they felt that some explicit coordination was required. On the other hand, Nickerson and Skiena [59] suggest a "call center" approach which ends up being quite similar to our setup, though with slightly more coordination on the operator side.

3.1.2 Number of Tasks

Some tasks benefit from being divided up and performed in parallel, while others do not. As discussed in Chapter 2, Steiner [75] calls these unitary and divisible tasks. He explains that unitary tasks will not significantly benefit from additional people working on the task. At best, unitary tasks are additive, meaning that the work done by individuals is cumulative and grows linearly with the number of group members. At worst, communication overhead and other issues related to the complexity of teams will negatively impact performance, resulting in poorer performance as more team members are added.

Given that tasks may or may not benefit from being divided up and performed in parallel, the benefit from assigning tasks to groups manifests itself in situations where a true division of labor is possible [75].

To that end, we tested the capabilities of multiple operators by using an apparently additive unitary task as well as a divisible task that can be divided into subtasks as defined by Steiner's taxonomy [75]. Since divisible tasks have the potential to show greater than linear growth in productivity as the number of group members increases, we hoped to see such improvement in the performance of operator teams in divisible tasks.

3.2 Experiment Design

Our experiment was designed to exploit the nature of group dynamics to see if greater performance could be gained from having operator teams work together on certain types of tasks. At the same time, we hoped to test the validity of Steiner's framework, meaning that we expected to see little or no improvement using teams in the case of an apparent unitary task, while seeing greater than linear growth in performance measurements in the case of a divisible task.

3.2.1 Test Scenario

The testing scenario simulated an exercise in which subjects took on the role of explosive ordinance disposal (EOD) workers. Using a simulation in which they had robots equipped with explosives-detecting sensors, the subjects had to search through a set of buildings searching for explosive devices. There were two different types of robots, each designed to find different types of explosives. "Bomb-sniffing" robots searched for bombs inside of buildings by following the "chemical plume" surrounding the bombs. "Mine-sweeping" robots were used to search for mines buried outdoors by sweeping an operator-designated search area. This was a 2x2 experiment design (see Table 3.2.1 with the number of operators and the number of tasks as the independent variables. Each test subject participated in only one condition, making this a between-subjects design. We

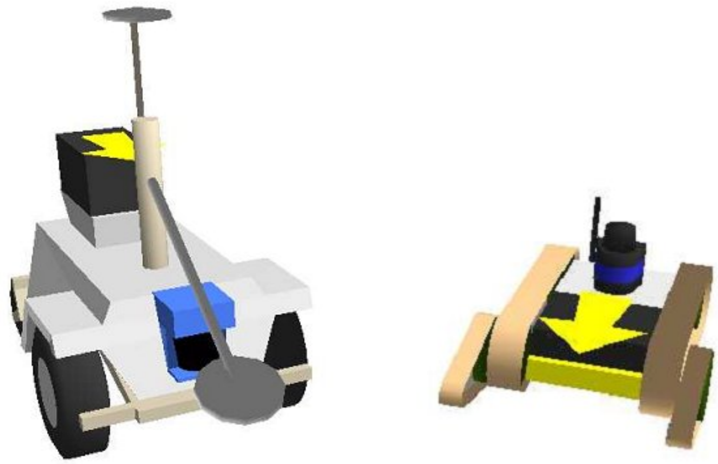

Figure 3.1: Up close view of the robot 3D models.

compared the results of participants working on their own and two participants working together in teams.

Due to the different way in which the “bomb-sniffing” and “mine-sweeping” robots are operated, they constitute two logically separate tasks, with a limited amount of overlap. The mine-sweeping robots generally require much less input since they are automated while searching for mines, whereas the bomb-sniffers require significant operator control while searching for a bomb since there is no automated behavior for support.

We designed these differences to create situations in which we could compare (a) teams performing (what we presumed to be) a unitary task to (b) a divisible task in which operators could assign themselves specific, complementary roles. The complementary nature of mine-sweeping and bomb-sniffing was increased by placing mines only on the exterior of buildings, and bombs only on the interior. This forced operators to clear the mines from entrances to buildings before they could enter to look for bombs.

Mines were cleared by having a mine-sweeper pass over them while in search mode (a mode initiated by the user). Bombs were cleared by a user placing an icon on top of the bombs. In order to find the bomb’s location, the operator had to follow a colored gradient, shown in Figure 3.2.

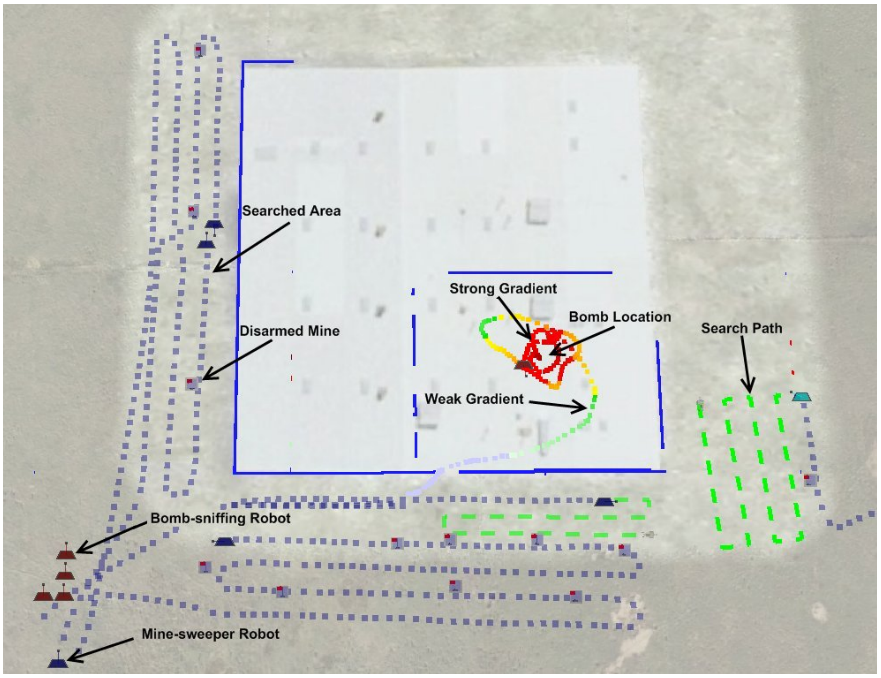

Figure 3.2: When zoomed out, robots appear as red or blue polygons, depending on the type of robot. The areas searched by the mine-sweepers are represented by the blue squares on the map, and the readings from the bomb-sniffing robot are shown as a colored gradient, going from green to yellow to red, getting darker when getting closer to the bomb.

3.2.2 Participants

A total of 64 paid subjects participated in the study. Of those, 16 were not included in the final results because of either (a) technical problems (interface or simulator locking up or having other issues) or (b) having their data invalidated by changes to the experiment design early on (in the case of a few early participants the bombs-only condition was modified slightly part-way into the study to include more bombs, invalidating data collected under the earlier conditions that has fewer bombs).

Of the remaining 48 individuals, there were 28 male and 20 female participants. Their ages ranged from 19 to 27, with a median age of 22. Of the 16 teams, 9 were married couples and the remaining 7 were friends who signed up together (3 were all-male, 2 were all-female, and 2 were male/female pairings).

Experiment Conditions	
One Operator Bombs-Only	Two Operators Bombs-Only
One Operator Bombs-and-Mines	Two Operators Bombs-and-Mines

Table 3.2: Experiment conditions

41 out of 48 participants responded that they were “Experienced” or “Very Experienced” with using a keyboard and mouse, and 23 out of 48 reported being “Experienced” or “Very Experienced” with playing video games.

3.3 Software Environment

In an effort to increase the real-world applicability of our results, we chose tasks that are very similar to those currently being researched at the Idaho National Lab (INL). Because of this similarity, we could use the augmented virtuality interface developed at the INL [8]. This allowed us to leverage the framework in the interface for visualizing and controlling multiple robots and their sensor readings.

A back-end to the interface was also required to simulate multiple robots and their interactions with the environment. Since no existing software met the requirements imposed by the experiment design and interface, we opted to create our own robot simulator. It supports dynamically created maps which are explored by “range sensors” on the robots, waypoint following and automated path planning. The simulator also acts as a server and coordinates communication between the interfaces when there are multiple operators. For design and implementation details, see Appendix 5.2.

3.3.1 Maps

Two different maps were used for testing and one map was used for training. Each map was used an equal number of times for each condition (four times). The participants were only told how many buildings there were, and that there could be more than one bomb per building. They were not told how many bombs or mines there were to find, just that the experiment would end once they found them all or time ran out (they were told the entire experiment would last no more than 75 minutes, including training time). Figure 3.3 shows the two maps used from both the user interface view (only a satellite image) and the simulator view (shows floor plans for each building, as well as positions of bombs and mines).

Each robot had a virtual sensor which detected obstacles in front of it. When an obstacle was detected, it was automatically added to a global map shared by all robots, and displayed in the interface as bright blue lines. Figure 3.2 shows the outlines of a building being discovered by robots and dynamically updated in the interface.

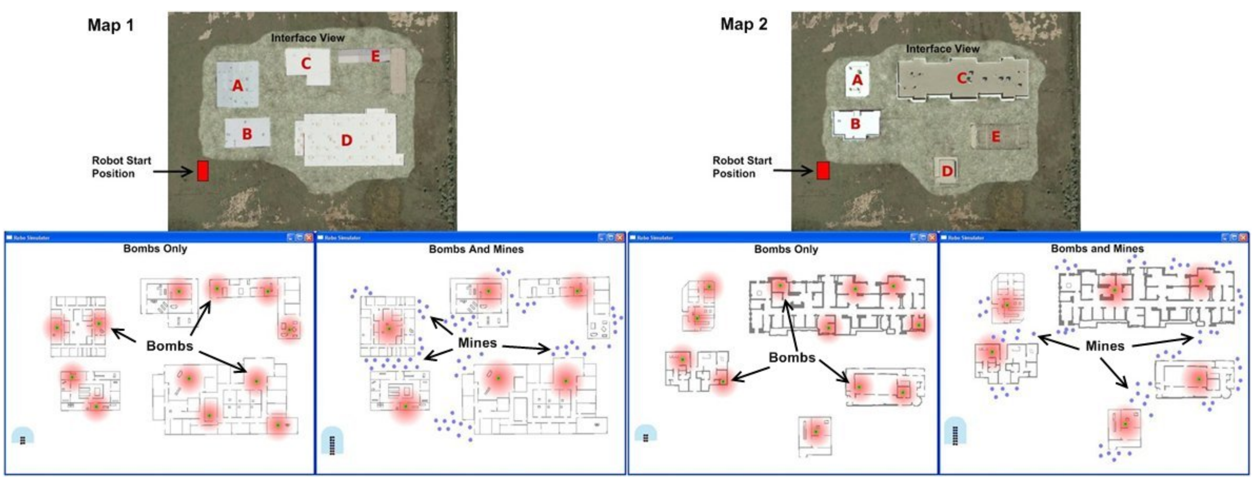

Figure 3.3: The maps used in the experiment. The red areas represent the chemical plume put out by the bombs which is detected by the bomb-sniffing robots, the blue dots represent mines, and the blue area immediately around the robot represents their scanner range in which obstacles can be detected.

3.3.2 Robot Autonomy

To support some degree of fan-out, we provided some limited autonomy to the robots to assist operators in performing their task. Without this basic autonomy, Neglect time would be close to zero, meaning that an operator would be consumed with controlling just one robot. Some level of autonomy is required in order to allow operators to spend a significant portion of their time making higher-level cognitive and planning decisions, which is where we believe the benefit of teams can be manifested.

To that end, each robot had the ability to follow paths set by the operator manually or to automatically generate their own path to a user-specified go-to point. Manually created paths were generally shorter and more direct than automatic paths, but could not be responsive to new obstacles being revealed in a dynamic environment. Automatic paths were less efficient, but would automatically re-plan to avoid obstacles discovered en route. Users also had the option of manually controlling the robots through the use of arrow keys on the keyboard. These trade-offs created an environment where users had to decide for themselves what level of autonomy was appropriate for a given situation (called adjustable autonomy [18]), requiring frequent planning and decision making.

Additionally, mine-sweeping robots had a "search mode" in which the user could indicate an area to be searched for mines, and the robot would create a search path that attempted to fit the area the user specified. In search mode, just as in following a manually created path, robots could easily get stuck on obstacles. This was a significant source of frustration for participants, since they could not trust mine-sweepers to complete a search area without becoming stuck. Also, the search path generated by the robots often did not match what the user intended, forcing them to recreate the search area until they were satisfied. Bomb-sniffing robots had no equivalent behavior to the mine-sweeper's search mode.

Another significant difference between mine-sweepers and bomb-sniffers was that mine-sweepers discovered and disarmed mines automatically while in search mode, while bomb-sniffers merely indicated how close the robot was to the bomb in a "warmer/colder" fashion (see Figure

3.2). This meant that operators had to pay much closer attention to robots getting close to a bomb, and most chose to drive them manually at the point. Any robot that drove over an explosive (mine or bomb) exploded if the explosive was not disarmed, forcing operators to use caution.

3.3.3 Communication

In addition to the primary task of finding explosives, we created a secondary task in which participants had to respond to a "commander" who would ask them questions about the environment such as how many explosives had been found, how many robots had exploded, how long they had been searching, and so on. These questions were picked at random from a predefined list, and would appear in a chat window located at the bottom left of the interface (see Figure 3.4) every 12-13 seconds. When a question was not answered, it was repeated instead of asking a new question. The accuracy of the answers was also recorded.

Between questions, meaningless messages would appear coming from "channel chatter". Four to five "distractor" messages were sent between every question from the commander. These forced participants to monitor the chat window more closely for relevant communication.

Teams also used this chat window to communicate. Participants were told not to talk, gesture, or communicate in any way except through the chat window provided. Because of this, those in teams not only had to monitor the window for commander questions, but for messages from their teammate as well. Distractor, teammate, and commander messages were each displayed in a unique color to give some assistance in distinguishing between them.

3.3.4 Scoring

As a method of simple motivation, participants were shown a running score that was based on their performance. Positive points were awarded for successfully disarming bombs and mines. Points were taken away for 1) losing a robot, 2) incorrectly marking a bomb, and 3) the length of time taken to complete the task (one point per second). The score was not affected by participants' answers to questions from the "commander", though in hindsight this may have been useful.

Figure 3.4: The interface with the chat window showing in the bottom left.

The point of scoring was to have a simple incentive to make participants challenge themselves and take the task more seriously. Even though the score was mostly meaningless and mentioned only briefly during the training, after completing the experiment nearly all the participants asked how their score was in comparison to others that had done it. This indicates that they were at least conscious of their score (and by extension, their overall performance) throughout the experiment.

Incentivising performance has implications for teams as well. Zaccaro et al. [85] found that giving an incentive for team performance in a timed situation did in fact improve team performance and coordination. Giving a score was a simple method of achieving at least a portion of this effect and preventing apathy in the test subjects.

3.3.5 Training

In order to help prevent a learning effect from confounding the data, each subject completed a short training course (20-30 minutes). The course consisted of a series of on-screen videos which explained the tasks, the interface, and the controls they would use to complete the tasks. They were given two to three opportunities to practice what was explained in the videos, depending on what condition they were participating in (those using only bomb-sniffers did not have the opportunity to practice with mine-sweeper robots, for example). Teams received an additional practice of three to four minutes to experience how coordination and communication would be handled (the rest of the training was conducted individually).

Subjects practiced on a different, smaller map than they ran the experiment on (see Figure 3.5).

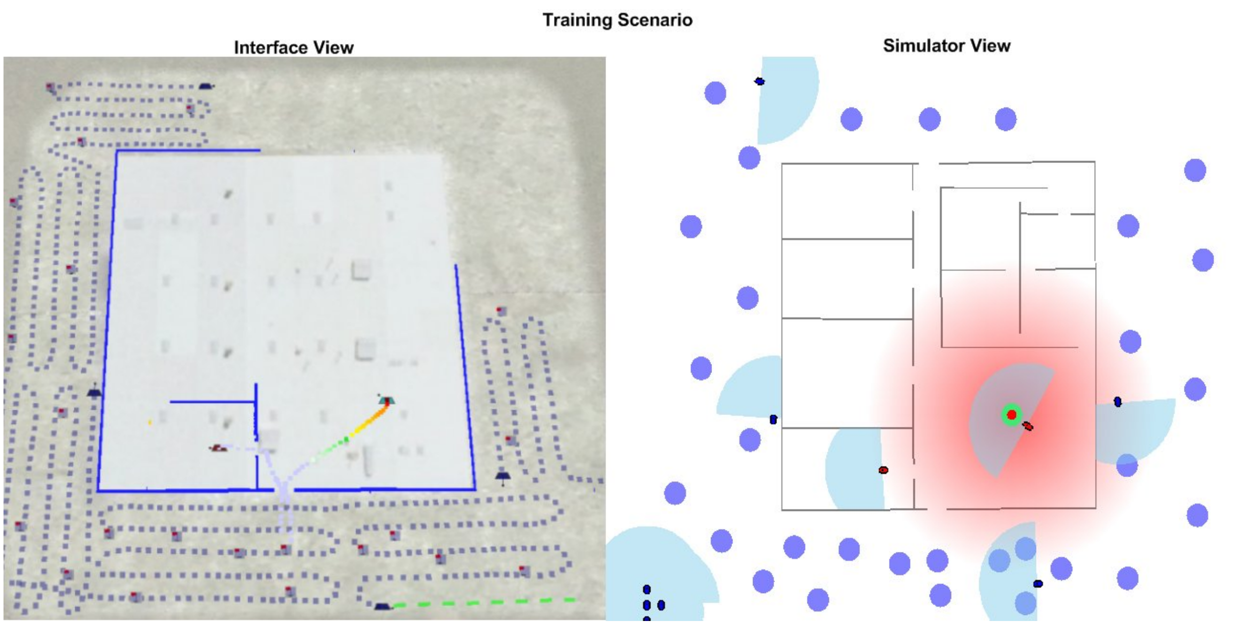

Figure 3.5: The map used for training and practice.

3.4 Dependent Variables

In this experiment, we expected to see improvement in two basic areas: task performance and workload. In order to show an overall benefit to using two-person teams over an individual, we expected the results to show a greater than linear improvement gained by using teams. Anything less would have indicated that there is little or no benefit to having operators work together. In order to look at task performance, we measured many different variables including the number of

Dependent Variables	
Bombs Found	Bomb-sniffer Fan-out
Mines Found	Mine-sweeper Fan-out
Robot Utilization	Total Fan-out
Unused Robots	Commands sent to Robots

Table 3.3: Dependent variables measured in the experiment.

"Commander" Questions
"How many mines have been found?"
"How many bombs have been found?"
"How many robots have exploded?"
"How many robots are currently following a path, including search areas?"
"How many bomb icons have been placed incorrectly?"
"How many robots are not being used right now?"
"How many minutes have you been searching the area?"
"What is your current score?"

Table 3.4: Questions asked by the "commander" in 12 second intervals. Questions about mines were only asked when mines were actually present.

bombs found, the number of mines found, the amount of time robots were utilized, and several others (see Table 3.3).

Workload was measured indirectly by measuring subjects' performance on a secondary task (answering questions from the commander), and by a subjective survey based on NASA TLX [45] (see Appendix B).

Chapter 4

Experimental Analysis

Once we began to analyze the data, we were faced with how to make fair comparisons between individual operators and teams of operators. In order to make it worthwhile to consider using operator teams, the results need to show greater than linear growth in effectiveness as operators are added to the team. For our experiment, that implies that teams of two operators would need to be more than twice as effective as an individual working alone. For statistical purposes, we measured individuals' performance separate from their teammates. Measuring in this way allowed for simpler comparisons between individuals working alone and individuals working as part of a team.

An important factor to consider is the added obstacles that teams have to overcome in order to be productive when compared with individuals. The costs of participating in a group (outlined in Chapter 2) can be a significant drain on the cognitive resources of operators, effectively acting as a hidden task. The requirements these costs placed on operators can be collectively thought of as a *coordination task* that only those operators in teams were required to perform.

A confounding factor was that the task was significant enough that very few subjects completed the entire experiment (i.e., found all of the explosives) because we told them we would not keep them beyond 75 minutes. Subjects were allowed to practice during the training until they decided that they were comfortable with the interface and tasks, which resulted in variable training time. These factors taken together mean that no two experimental runs can be compared side by side. Taking the number of explosives found in one experimental run cannot be compared to the number from another run when the running time of the experiment is very different.

To resolve this issue, we limited our examination to the first 20 minutes of each experimental run. Since the shortest completion time was slightly more than 20 minutes, this gave us equal footing to compare performance. For statistical analysis of means, we used the results recorded at the 20 minute mark; however we corroborated those results by looking at trends over a the entire 20 minute time period. All significance values (p-values) reported were calculated using an independent two-tailed T-test.

4.1 Primary Task Performance

In this section we present the results of the primary task(s) and discuss other measures related to this performance. Unless otherwise noted, results are presented at a per-operator level for comparison purposes, meaning that the total performance of teams is twice what is listed in the tables and charts below.

4.1.1 Bombs Only

In the *bombs-only* condition, both individual operators and operator teams had twelve bombs to find in one of two similar maps, each of which contained five buildings.

In this condition, there was only a single task which required considerable attention from operators, because the process of searching an area for bombs could not be automated like it could with mines. Thus, we expected that having a team of operators would show at best linear improvement over a single operator. Under Steiner's framework [75], the bombs-only condition is called an *additive unitary* task.

This turned out to be a fairly accurate prediction. Figure 4.1 shows the cumulative total of bombs found over a 20 minute period. Eight operators working alone found an average of 5 bombs, while operators working in teams found an average of 3.56 bombs per operator. The difference was not statistically significant (see Table 4.1 for significance values). In fact, very few of the measures were significant between single operators and operators on teams, which corresponds with how Steiner defined an additive unitary task.

Notable Results for One Task			
Measure (per operator)	**Operators Working Alone**	**Operators on Teams**	**Significance (p-values)**
Bombs Found	5.00	3.56	p = .192
Fan-Out	3.47	2.19	p = .0004
Total Commands Sent To Bomb-sniffers	1819	1103	p = .018

Table 4.1: The most significant results from operators working alone and as teams for the bombs-only condition.

These results show linear growth for primary task performance using teams of operators. Fan-out, however, dropped by about one-third when operators were in teams, meaning there was not completely linear improvement with operator teams. We will further discuss fan-out in a later section.

4.1.2 Bombs and Mines

Following Steiner's framework, we also tested a divisible task where operators had to search for *bombs-and-mines*, with each type of search requiring different approaches. Searching for bombs is more of a manual task requiring close supervision of the robots, while searching for mines is more or less automatic once a search area is designated. Our setup follows Steiner's definition of a divisible task.

Operators had 6 bomb-sniffer robots and 8 mine-sweeper robots at their disposal to complete this task. This meant that there were 14 robots to monitor and command compared to only 6 in the bombs-only condition. Operators searched 5 buildings to find 6 bombs, and searched the area around those buildings to try to find 65 hidden mines. Based on subjective and objective workload evaluations, this condition was significantly more taxing than the bombs-only condition.

The performance of operator teams for finding bombs grew linearly. Individuals working alone averaged 1.25 bombs while those working in teams found 1.44 bombs per person. The difference was not statistically significant (see Table 4.2), however there was a strong trend toward greater than linear growth (see Figure 4.2).

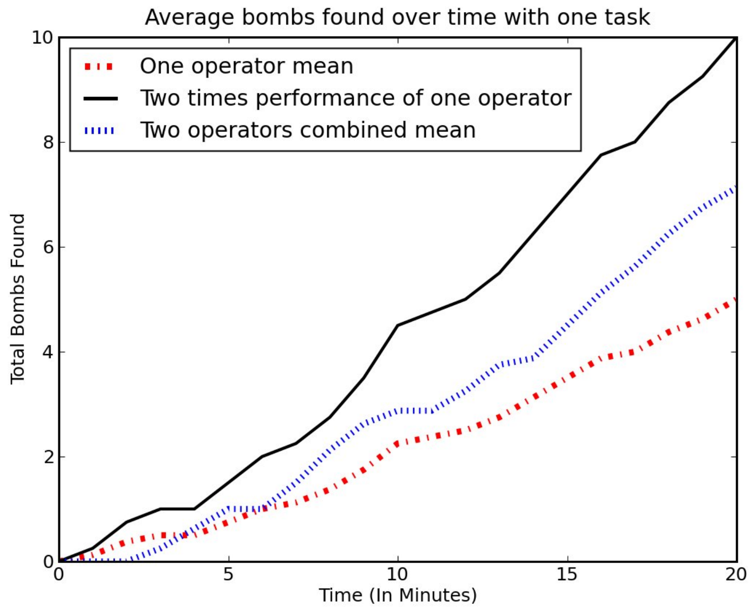

Figure 4.1: Cumulative total of bombs found over a 20 minute period in the bombs-only condition. Team results are reported as the average both team members' individual results.

We were somewhat surprised by the teams' performance in finding mines. The average number of mines fell from 28 per individual working alone to 15.25 per individual working on a team. The cumulative trend is shown in Figure 4.3. We we would have expected the number of mines found to grow at least linearly, just as the number of bombs found did. It was especially surprising since mine-sweepers are heavily automated and therefore have a larger neglect time when compared to bomb-sniffers. These factors should have made the task easier, not more difficult.

This discrepancy might be explained by two factors. First, in order to search for mines a significant amount of ground had to be covered by the mine-sweepers. Thus, performance was severely limited by physical constraints (e.g., travel time) and not cognitive constraints (e.g., time taken to plan an optimal route to search a building for a bomb). Many mines were located on the opposite side of the map from where the robots started, a large distance from the robots' start position. We believe that if we would have more evenly distributed the robots across the map to start with, that travel time would have been mitigated, and the number of mines found by teams would have increased at least linearly.

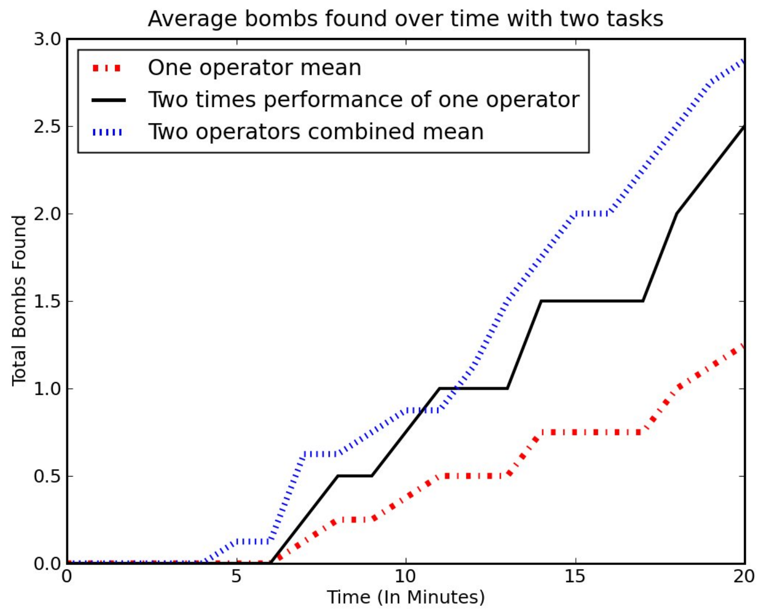

Figure 4.2: Cumulative total of bombs found over a 20 minute period in the bombs-and-mines condition. Team results are reported as the average sum of both team members' individual results.

Second, there were some mines that were much more difficult to find than the rest. The mines were primarily clustered around entrances to the buildings (which the subjects were told), but a portion of them were hidden elsewhere. In addition, some of the entrances were on the far side of buildings, and there was no need to clear them for the bomb-sniffing robots to get inside; This made searching those areas a low priority. In essence, the more mines that were found, the more difficult the task became.

These explanations are supported by looking at individual results. Within the operator teams, it was common (but not universal) for one member to focus on searching for mines, while the other team member focused on searching inside of buildings for bombs. If we look at the number of mines found by individuals in a team, and then separate the operators into two groups based on which operator found more mines than their teammate, the average number of mines found by the "specialist" teammates is 26.13, compared to an average of 4.38 for the "non-specialist" teammates. This number is much closer to the number of mines found by operators working alone, and

Notable Results for Two Tasks			
Measure	**Operators Working Alone**	**Operators on Teams**	**Significance (p-values)**
Bombs Found	1.25	1.44	p = .665
Mines Found	28.00	15.25	p = .04
Fan Out for Bomb-sniffers	1.95	1.33	p = .1
Fan Out for Mine-sweepers	4.77	2.82	p = .03
Total Fan Out	6.71	4.14	p = .015
Total Commands Sent To Bomb-sniffers	857	865	p = .97
Total Commands Sent To Mine-sweepers	4382	2378	p = .052
Unused Bomb-sniffers	1.0	.25	p = .059
Unused Mine-sweepers	.375	.125	p = .49
Waypoints given to Mine-sweepers	4303	2317	p = .051

Table 4.2: The most significant results from operators working alone and as teams for the bombs-and-mines condition.

is consistent with the explanation that there was a performance "cap" for mine-sweeping robots due to the experimental setup.

4.1.3 Summary

These results show that for small operator teams, there appears to be linear growth in performance. The number of explosives found grew approximately linearly for both the bombs-only and bombs-and-mines conditions. This means that while there was no benefit on the primary task(s) to having operator teams, there was little or no cost associated with the "coordination task" the teams had to deal with.

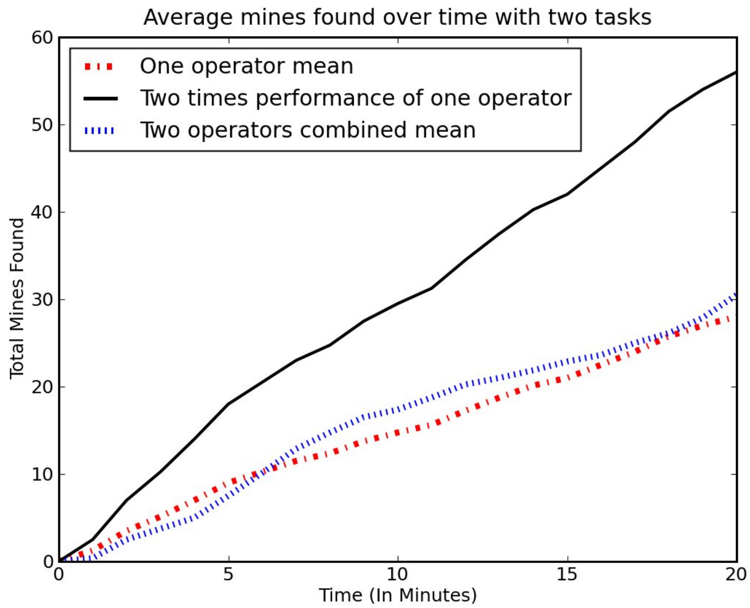

Figure 4.3: Cumulative total of mines found over a 20 minute period in the bombs-and-mines condition. Team results are reported as twice the reported value in Table 4.2.

4.2 Workload

Workload was measured in two ways: through performance on a secondary task of answering questions posed by a "commander" through a chat window, and a subjective survey based on NASA TLX [45] given at the end of the experiment. These are both commonly used methods for measuring for workload [78]. The full text of the subjective workload survey is in Appendix B. The accuracy of the answers was measured as well, but was found to have no significant different between experiment conditions.

4.2.1 Bombs Only

As shown in Table 4.3, in the case of both secondary task performance and the subjective survey results, there is no statistically significant difference between individuals who worked alone and individuals working on teams. The linear growth trend is especially apparent in Figure 4.4. This again validates what we expected to see for the bombs-only condition. It reflects the capacity for linear growth in performance for teams, with the potential for only small overhead due to the hidden coordination task.

Workload Measures for One Task			
Measure	**Operators Working Alone**	**Operators on Teams**	**Significance (p-values)**
Questions Answered	29.75	30.31	p = .92
Subjective Workload Measures			
Mental Activity	3.25	2.88	p = .461
Time Pressure Felt	3.25	2.88	p = .409
Mental Work Required	3.0	2.75	p = .56
How Successful	3.13	3.07	p = .89
How Satisfied	3.38	3.07	p = .52
How Relaxed	3.0	3.06	p = .871

Table 4.3: Results from workload assessments of operators working alone and as teams for the bombs-only condition.

4.2.2 Bombs and Mines

Table 4.4 summarizes the workload measures for operators working on two tasks. These results are striking when compared to the primary task performance of the two groups. Individuals operating together answered 42% more questions from the "commander" than operators working alone (see also Figure 4.5). Additionally, reported mental activity and mental work fell by 36% and 20% respectively, and time pressure felt was reduced by 27%. Again according to the subjective survey, 75% of subjects working alone thought they were either "Unsuccessful" or "Somewhat Unsuccessful" in accomplishing the tasks while nearly 69% of subjects working together reported at least "Satisfactory" success.

4.2.3 Summary

There was a significant increase in workload when operators were asked to complete a second task (see table 4.5). These differences in workload completely disappeared when the subjects were

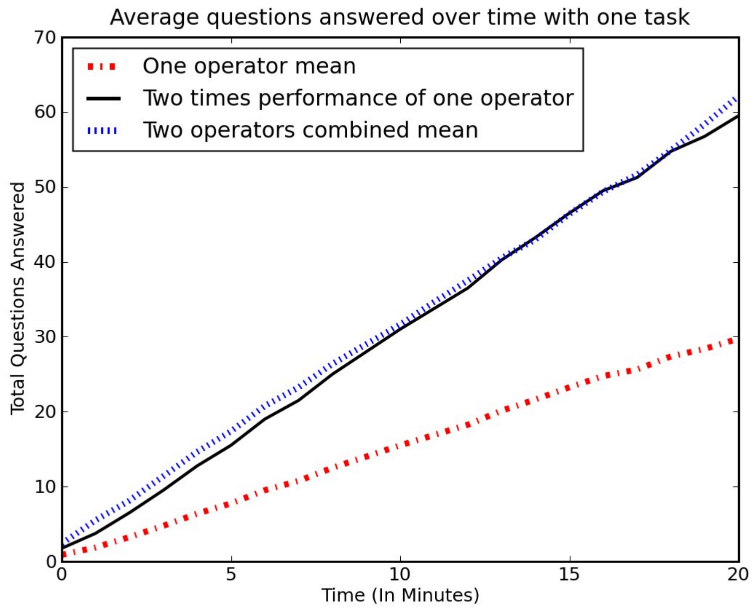

Figure 4.4: Cumulative total of questions answered in the bombs-only condition. Team results are reported as twice the reported value in Table 4.3.

working in teams. Statistically, there is no significant difference in the workload measures for operators working on one task alone and for operators working on two tasks as part of a team.

4.3 Fan-out

Another point of interest is the fan-out metric, with the results shown in Tables 4.1 and 4.2. We measured fan-out by counting the number of robots actively being controlled by each individual at 30 second intervals, and then averaged over the first 20 minutes of experiment time. We defined an active robot as one that moved within the last second. It is important to note that our direct measurement of fan-out for the heterogeneous bombs-and-mines condition differs from the model in Chapter 2, which indirectly predicts the maximum fan-out for homogeneous robots performing identical tasks.

Workload Measures for Two Tasks			
Measure	**Operators Working Alone**	**Operators on Teams**	**Significance (p-values)**
Questions Answered	19.88	28.38	p = .042
Subjective Workload Measures			
Mental Activity Required	4.63	2.94	p = .0009
Time Pressure Felt	4.25	3.06	p = .0126
Mental Work Required	3.63	2.88	p = .052
How Successful	2.0	2.94	p = .037
How Satisfied	2.38	3.0	p = .229
How Relaxed	2.13	2.75	p = .124

Table 4.4: Results from workload assessments of operators working alone and as teams for the bombs-and-mines condition.

4.3.1 Fan-out Results

The fan-out measures show a decline of more than 35% in the number of robots controlled by individuals that worked in teams compared to those who worked alone (for both the bombs-only condition and the bombs-and-mines condition). However, as our data shows, a decrease in fan-out does not always correspond to a decrease in performance.

It is important to point out that overall fan-out for teams did not actually decrease, it just did not grow linearly (see Figure 4.6). However, the total number of active robots was still well below the maximum possible. For the bombs-only condition (6 robots available), individuals had an average fan-out of 3.47 and teams had a combined average fan-out of 4.38. For the bombs-and-mines condition (14 robots available) individuals averaged a fan-out of 6.71 and teams averaged 8.28.

It is not apparent from the data whether the lower individual fan-out contributed to lower workload for operator teams, but we suspect this is the case. If this is true, team members had additional capacity and were limited by the constraints the experiment imposed (e.g., travel time for the robots). Future experiments might be able to take advantage of this "spare capacity" and

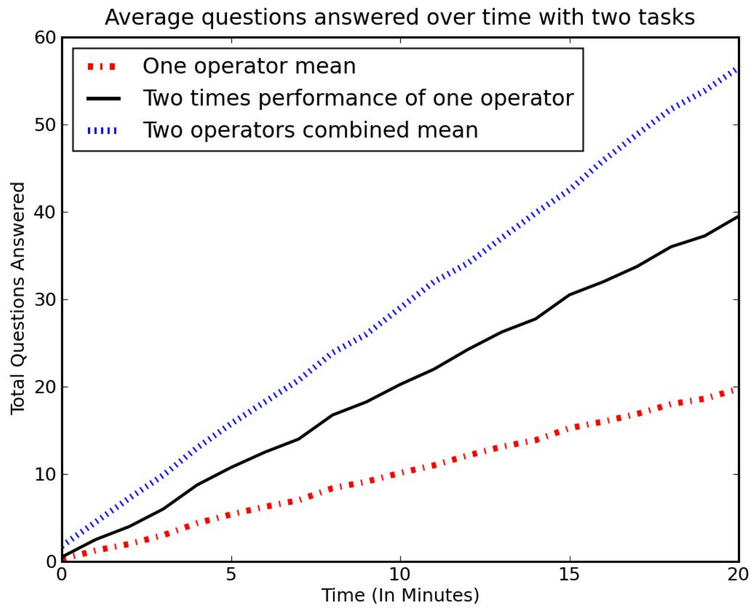

Figure 4.5: Cumulative total of questions answered in the bombs-and-mines condition. Team results are reported as the average of the sum of both team members' individual results.

allow operators to increase their performance further, thus definitively showing greater than linear growth in performance on primary tasks for operator teams, rather than just workload reduction.

4.3.2 Fan-out and Performance

One trouble with focusing on fan-out is that increasing the number of robots controlled does not guarantee improved performance. Olsen and Wood give two reasons for this: *task saturation* and the *fan-out plateau* [63]. Task saturation occurs when a task will not benefit from having more robots assigned to it. A simple example is a task to transport a small item from one location to another. Two robots would not allow the task to be completed any faster than one would. The fan-out plateau is a point at which increasing fan-out will not improve performance due to less effective utilization of individual robots.

Task saturation explains the disparity between the upward trend for fan-out in bomb-sniffer robots shown in Figure 4.6(B) and the relatively steady rate of finding bombs shown in Figure 4.2 under the bombs-and-mines condition. Mine-sweepers demonstrate this trend as well in Figures

Workload measures for Operators Working Alone			
Measure	**Operators Performing One Task**	**Operators Performing Two Tasks**	**Significance (p-values)**
Questions Answered	29.75	19.88	p = .128
Subjective Workload Measures			
Mental Activity Required	3.25	4.63	p = .020
Time Pressure Felt	3.25	4.25	p = .040
Mental Work Required	3.0	3.63	p = .230
How Successful	3.13	2.0	p = .100
How Satisfied	3.38	2.38	p = .131
How Relaxed	3.0	3.625	p = .045

Table 4.5: Results from workload assessments of operators working alone and as teams for the bombs-and-mines condition.

4.2 and 4.6(C). The graphs show that although an average of 20% more mine-sweeper robots were being used, the number of mines found only increased by 8.9%. In the absence of task saturation, we would expect to see at least a linear improvement with the number of robots. Since there were two operators present, this discrepancy is not explained by the fan-out plateau.

The graph also shows a cycle for the bomb-sniffing robots related to fan-out. Bomb-sniffers differed from mine-sweepers in that they did not have an automated behavior that could be used to search for explosives. Because of this, mine-sweepers could be effective in finding explosives without continuous operator direction, while bomb-sniffers could not. This led to a common operating strategy to send a few bomb-sniffers on a path to an unexplored location so that the operator could focus on manually controlling just one bomb-sniffer. Dips in the fan-out graphs show the points at which robots were arriving at their destination before the operator was done manually searching (shown in Figure 4.7). Once finished searching, the operator would again give bomb-sniffers new paths and select one robot to focus on for manual control, and the cycle would start again. Three of these cycles are apparent from Figure 4.7 for operators working alone. These cycles occur with operator teams as well, but are less apparent, probably due to operators sharing task load.

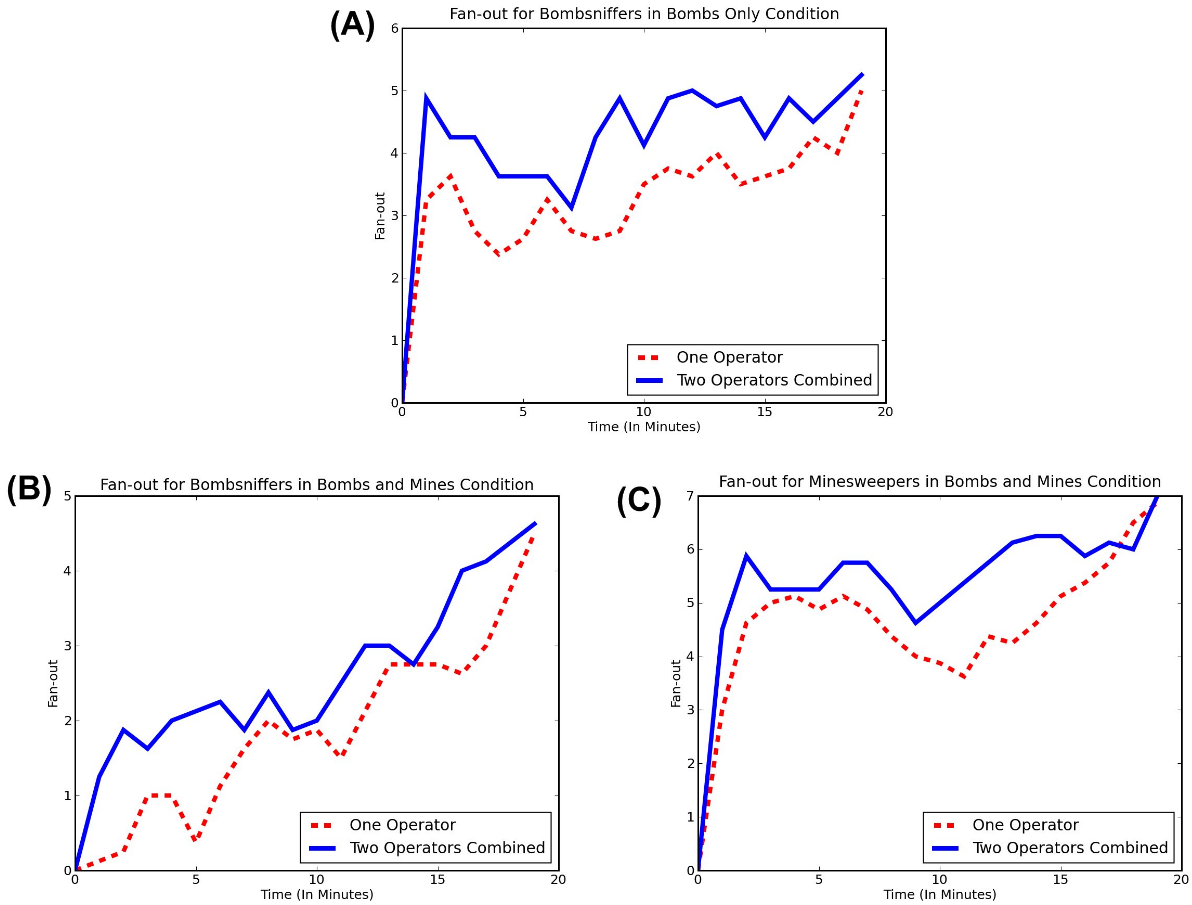

Figure 4.6: (A) Fan-out for bomb-sniffers in the bombs-only condition, (B) Fan-out for bomb-sniffers in the bombs-and-mines condition, and (C) Fan-out for mine-sweepers in the bombs-and-mines condition. Team results are shown as combined totals from both team members.

Figure 4.8 shows the fan-out for bomb-sniffers and the number of bombs found over time. There is a stair-step appearance to the graph of bombs found over time, with performance plateaus where no bombs were found for one or more minutes. This appears to be a period of time in which operators were not able to effectively search for bombs due to waiting for travelling bomb-sniffers to arrive at new locations or for mines to be cleared near building entrances. Individual operators had longer plateaus, indicating that they were busy performing other tasks besides searching for bombs, presumably because they were searching for mines. These plateaus were shorter for teams compared to individuals.

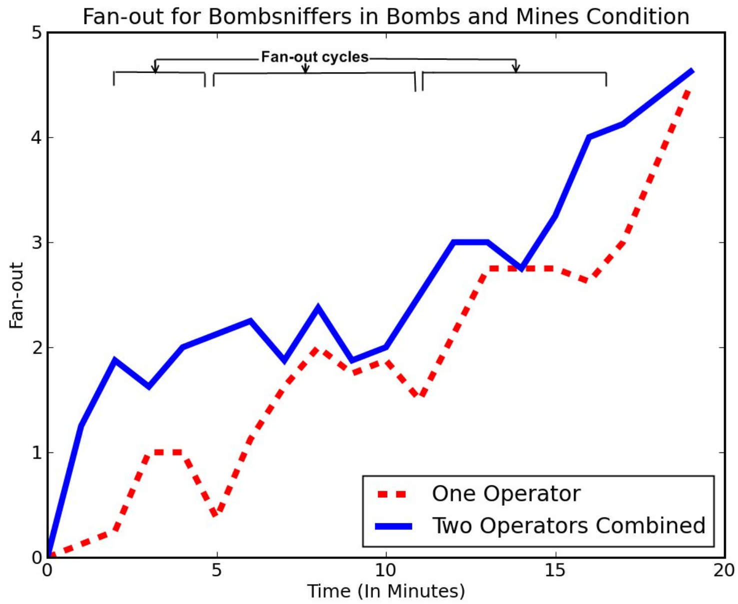

Figure 4.7: Cycles over time in bomb-sniffer fan-out in the bombs-only condition. Team results are shown as combined totals.

4.3.3 Summary

Task saturation appeared to play a significant role in the performance of operators and seems to explain the relationships observed between fan-out and performance. Operator teams did have a higher combined fan-out than operators working alone, but individual team members' fan-out was lower. This was probably caused by the constraints of the task environment rather than the effects of the hidden coordination task.

Regardless of the causes of lower fan-out for the operator teams, the data shows that it did not negatively impact performance on the primary task(s), and that operators in teams experienced reduced workload and perhaps more spare capacity to increase fan-out if the task allowed. Thus, lower fan-out correlates with lower workload without decreasing performance.

4.4 Training

Training was an important part of our experiment. Training allowed subjects to know what would be expected of them and how to accomplish the tasks they were required to do. Due to inexperience

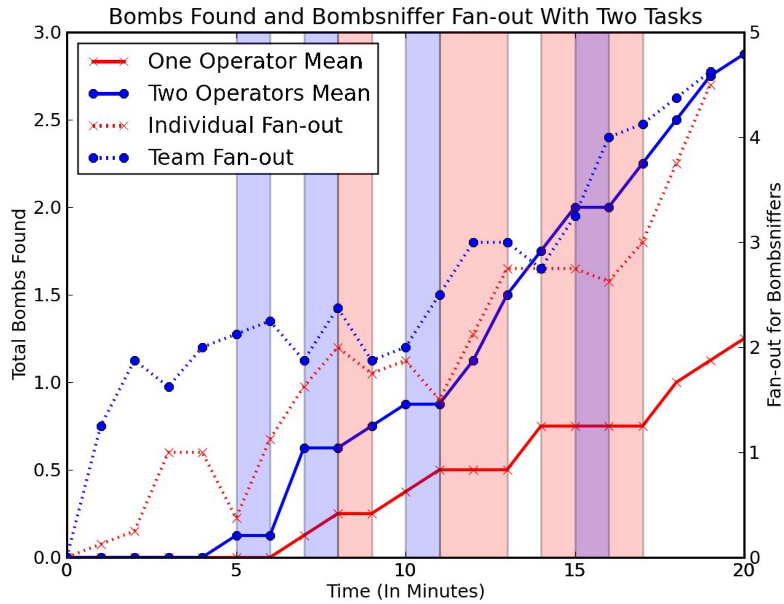

Figure 4.8: Bomb-sniffer fan-out and bombs found over time in the bombs-and-mines condition. Team results are shown as combined totals.

in performing the tasks in the experiment, unequal training can skew the results. This danger exists in our experiment, since we allowed users to practice as long as they felt comfortable, resulting in highly varied training times. Subjects that were part of a team also received additional training time to practice coordinating with their teammate.

Because of these factors, the average training time for operators working alone in the bombs-only condition was 7.17 minutes, while operators on teams averaged 12.65 minutes ($p = .007$). Operators working alone in the bombs-and-mines condition spent an average of 13.43 minutes in training, while operators on teams spent 19.41 minutes in training ($p = .012$).

This is important to look at since the additional practice time for those in teams has the potential to significantly bias the results. To see what biases, if any, training time had on performance in the experiment, we measured the correlation between training time and various performance metrics for every condition.

4.4.1 Bombs Only

Table 4.6 shows the correlation relationship between training time and performance for the bombs-only condition. As training time increased for individuals in the bombs-only condition, they found fewer bombs and answered more questions. This implies that more training leads to worse performance on the primary task, which seems illogical. A more likely explanation is that those who felt less comfortable with performing the task chose to practice for longer periods of time. Those who trained longer were probably less skilled at the task and the increased training time did not significantly help them. Those who were not as skilled either (a) focused more on answering questions, or (b) allowed the secondary task to interfere with their primary task performance.

Training Time Correlations for the Bombs-Only Condition	
Comparison	**Correlation Value**
Individual Operators	
Training Time vs. Bombs Found	-.426
Training Time vs. Questions Answered	.60
Operator Teams	
Training Time vs. Bombs Found	.099
Training Time vs. Questions Answered	.096

Table 4.6: Effects of training time on performance for the bombs-only condition.

For teams, this pattern disappeared. The correlation between training time and the number of bombs found, and training time and questions answered is very weak (.09). One explanation is that the increased training time for operator teams overcame the weaknesses for the less skilled operators, but the data does not support that conclusion. If we take the 8 operators who worked in teams with the most training time and compare their performance to the 8 operators who worked alone, we find that in spite of having more than twice the average training time of operators working alone (16.45 minutes vs. 7.17 minutes), the number of bombs found by those on teams drops to an average of 3, compared to 5 for operators working alone ($p = .145$).

4.4.2 Bombs and Mines

For the bombs-and-mines condition, there is a similar pattern for the effect of training time to that found in the bombs-only condition. Table 4.7 shows the correlations between training time and performance metrics for the bombs-and-mines condition. Here again, as training time increased for operators working alone, the number of bombs found went down, and the number of questions answered went up.

Training Time Correlations for the Bombs-and-Mines Condition	
Comparison	**Correlation Value**
Individual Operators	
Training Time vs. Bombs Found	-.727
Training Time vs. Mines Found	.251
Training Time vs. Questions Answered	.808
Operator Teams	
Training Time vs. Bombs Found	-.026
Training Time vs. Mines Found	-.240
Training Time vs. Questions Answered	.253

Table 4.7: Effects of training time on performance for the bombs-and-mines condition.

For teams, the trend again reversed itself. An increase in training time for operators on teams resulted in virtually no change for the number of bombs found, and an increase in the number of questions answered. None of these correlations are very strong for teams.

4.4.3 Summary

There are several reasons why the trends observed for individuals with regards to training time are not reflected for teams. Probably the most likely explanation is that there is just not enough data to establish a strong trend. With only 8 data points for individuals, just one or two outliers can form a trend. The increased number of data point for teams probably gives a more accurate picture of the relationship between training and performance.

If the trend is not just a statistical anomaly, then there are other interesting possibilities. One factor may be that operators on teams were able to specialize somewhat, allowing them to avoid their weaknesses and focus on tasks they were more effective with (even in the bombs-only condition, operators could focus on just answering questions). Another possibility is that operators on teams learn from each other, allowing less skilled individuals to see superior strategies from their teammate which they can then implement themselves.

In any case, the length of training time does not seem to be a factor in explaining the performance of teams compared to individual operators.

4.5 Environmental Factors

To help establish greater validity for our results, we randomly assigned participants to run the experiment using one of two different maps. The maps both had the same number of buildings, but the buildings had different floor plans and explosives were placed in different locations.

Performance comparisons for the different maps are shown in Table 4.8. The only statistically significant difference is that Map 2 appears more difficult for individual operators in the bombs-only condition. Since teams in that same condition showed almost identical results for the two maps, this is most likely a chance occurrence. None of the other conditions showed any significant differences. This implies that the map type had very little impact on the results of this experiment.

Metric	**Map 1**	**Map 2**	**Significance (p-values)**
Individuals With Bombs-Only			
Bombs Found	7.0	3.0	p = .049
Questions Answered	25.75	33.75	p = .468
Teams With Bombs-Only			
Bombs Found	3.5	3.6	p = .932
Questions Answered	28.16	31.60	p = .334
Individuals With Bombs-and-Mines			
Bombs Found	1.5	1.0	p = .866
Mines Found	24	32	p = .269
Questions Answered	18.75	21	p = .772
Teams With Bombs-and-Mines			
Bombs Found	1.5	1.375	p = .818
Mines Found	12.875	17.625	p = .545
Questions Answered	26.875	29.875	p = .512

Table 4.8: Performance comparison for different world maps.

4.6 Team Dynamics

An interesting aspect of this experiment comes when looking at how individual actions and performance differed between teammates. Understanding the dynamics of team interaction is important for any system designer looking to utilize operators teams. In this section, we discuss some observations from our experiment regarding these dynamics.

4.6.1 Division of Labor

The biggest reason that divisible tasks can benefit from groups is that they can be divided into logically separate subtasks that can be efficiently done in parallel. This is the benefit of a division of labor. By having two tasks, searching for bombs and searching for mines, we expected that most teams would have members which specialized into performing one task or the other. Test subjects were instructed that they would be more effective if they cooperated together, but were not assigned a specific role. In Steiner's framework, it is called *self-matching* when team members decide among themselves who will take on a given role. There were no definite roles given, and subjects were free to (and frequently did) switch roles from searching for bombs to searching for mines and back again. This most closely matches Steiner's description of *unspecified subtasks*.

In order to see how this division of labor took place and what its effect was, we correlated combined performance data for teams with the amount of specialization that took place within the team. We calculated a "specialization score" using the following formula:

$$\text{Specialization} = |\ \text{Teammate(A) Fan-out} - \text{Teammate(B) Fan-out}\ |$$

A higher value indicates that there was more specialization among the team (since there is a greater disparity in the number of robots of one type they controlled), and hence a greater division of labor occurring. These scores were calculated separately for bomb-sniffers and mine-sweepers. We also calculated a specialization score for questions answered, though this is not very insightful because team members were unable to coordinate performance on this task, and did not know what the other was doing in this regard (only the subject answering the commander's questions received any feedback).

Given these measurements, we found no correlation (coefficient = -.04) between bomb specialization and the number of bombs teams found (Figure 4.9), and a slightly negative correlation (coefficient = -.43) between question specialization and the number of questions teams answered (Figure 4.10). There was, however, a strong positive correlation (coefficient = .89) between mine specialization and the number of mines teams found (Figure 4.11).

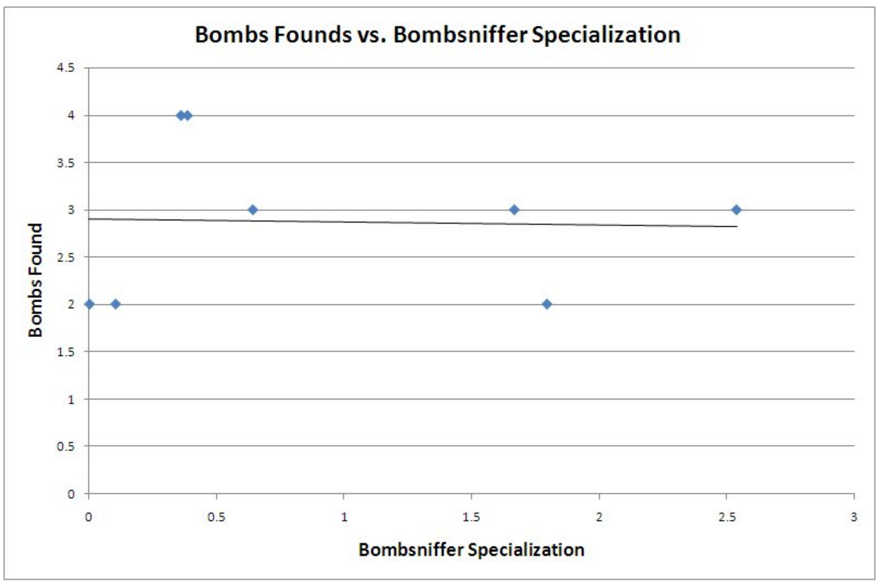

Figure 4.9: Scatter plot correlating bombs found with bomb-sniffer specialization with a correlation coefficient of -.04.

Specialization did not follow any clear trends that are discernible from this small of a data set. Prior to analyzing the data, we predicted that team members who focused primarily on minesweeper robots would also tend to answer more questions, since we supposed that directing the search areas for mine-sweepers has a lower workload than searching for bombs. This turned out to be the case in only one team out of eight. Other team members who answered more questions focused more on bomb-sniffing robots, or seemed to answer questions instead of working on the primary tasks, allowing their teammate to control more robots of both types. Several other teams had fairly even splits in terms of how many questions they answered and how many of each type of robot they controlled.

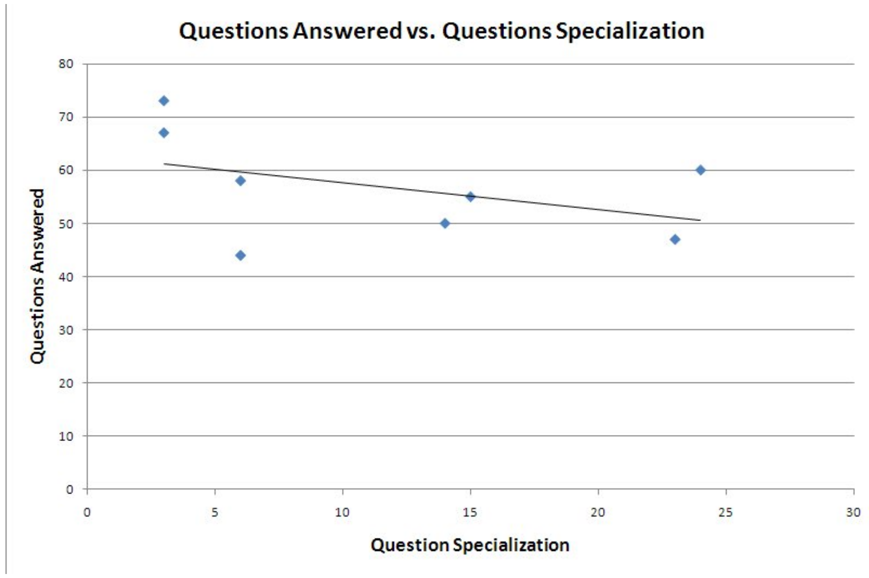

Figure 4.10: Scatter plot correlating questions answered with question specialization with a correlation coefficient of .89.

There is too little data to determine if one of these divisions of labor are more effective than others. This would be an interesting topic for future research.

4.6.2 Coordination

The primary mechanism we provided to communicate was a chat window that was open throughout the experiment. Though subjects were trained on using the chat window, few made very extensive use of it to communicate with their teammate. Subjects exchanged approximately 9 messages during the experiment, most of which were during the first few minutes. These initial conversations generally established a basic strategy for the tasks, as in this example between two subjects:

Subject A: so do you wanna divide bldgs

Subject B: do you want to do building b and I'll do A

Subject A: cool

Out of the 16 teams that participated in the experiment (8 performing one task, 8 performing two tasks), 11 had similar initial periods of coordination. Two of the teams in the bombs-only condition

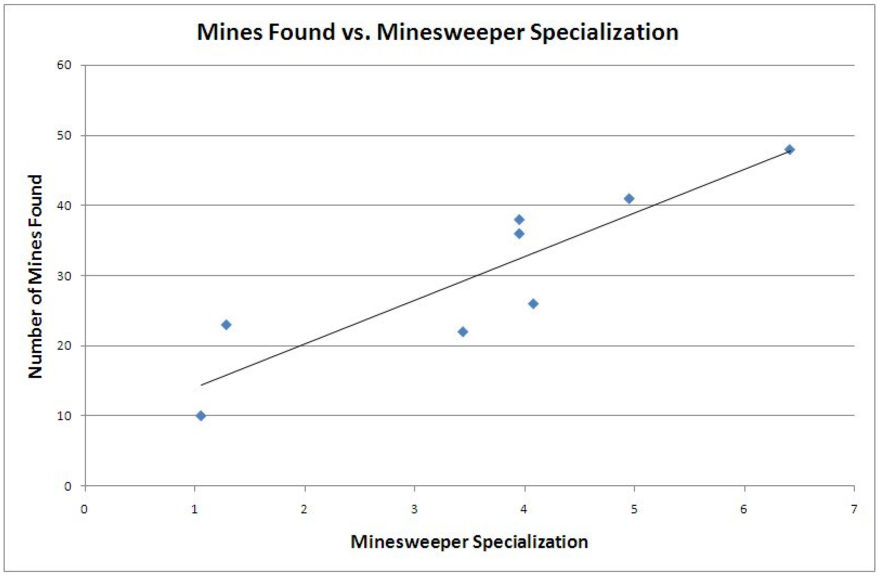

Figure 4.11: Scatter plot correlating mines found with mine-sweeper specialization with a correlation coefficient of .89.

did not communicate at all through chat messages, and two of the teams in the bombs-and-mines condition did not communicate initially, but did use the chat window later.

The relatively low volume of messages was likely due to the frequency of other messages being displayed in the same window, making it difficult for subjects to see when they were receiving messages from their teammate. Several subjects felt like their messages were not being received by their teammate at all, as in this one-sided conversation excerpt:

Subject: hello?

Subject: Hello?

Subject: HEY we need to work together can you see this?

A significant number of subjects complained after the experiment that there must have been some sort of technical difficulties preventing their messages from getting through. The messages were indeed sent, but were often missed due to the distractor messages and commander questions in the chat window.

In spite of these communication difficulties, teams seemed to have a fair bit of coordination taking place. Even though there was no consistent way in which tasks were divided among team members, most teams did exhibit a significant amount of specialization. This is somewhat remarkable due to the sparse communication that took place.

An explanation lies in the existence of what is called *implicit communication*. Implicit communication is where team members communicate without words (written or verbal) and can be a combination of body language, actions, and other factors. The existence of implicit communication can be a sign of effective team coordination, and has been found to increase performance if the shared understanding between team members is accurate [72].

Orasanu examined implicit communication in cockpit air crews, and found that team members will often alternate between implicit and explicit communication [64]. Explicit communication typically takes place in times of low-workload and often facilitates implicit communication later in times of high-workload. Explicit communication is often used to plan and coordinate future actions, or to communicate when an exceptional situation arises and implicit coordination is no longer sufficient.

This explanation from Orasanu fits our observations well. Teams generally coordinated at the beginning through explicit communication, and used implicit communication to coordinate otherwise. Occasionally subjects felt the need for explicit communication later on, usually for one of several reasons. First, subjects sometimes felt the need for a status "refresher" just to make sure they were still on the same page, by asking something such as "how are things going?" or by making a statement like "good job". Other times there was an expression of frustration resulting from some sort of misunderstanding such as "you're stealing my robots" or "hey you took my guy". This showed a lack of shared understanding between team members, prompting explicit communication. Finally, subjects would sometimes feel a need to coordinate again and lay more plans together, resulting in statements like "you finish building B" and "will you start searching building a?". This type of communication apparently took place when a member of the team felt like their effectiveness was in jeopardy and wanted to make plans once again.

An important aspect of implicit coordination is having team members give required information to teammates before it is explicitly requested, thus anticipating their teammates' needs [30].

This aspect of implicit communication is missing from our data. This is probably because 1) the task did not impose a situation where one operator was dependent on the other, reducing the need to share information and 2) any information that was needed was probably gained from the shared workspace provided by the interface.

We feel that the interface afforded significant implicit communication as well. The robot that was currently selected by a subject's teammate was shown on the interface, so they could know at all times what their teammate was doing. Subjects could also quickly see the effects of their teammate's actions, which is a form of implicit communication. The interface likely reduced the need for explicit communication and made implicit communication easier.

4.7 Confounding Factors

In spite of all the thought and planning that went into the experiment design, there are always factors beyond control or improvements that can be made with the benefit of hindsight. We discuss a few of those here.

Significant actions in the experiment were tied to a score in order to motivate participants to focus on the things we were measuring. Therefore finding bombs and mines, completing the experiment quickly, and not allowing robots to explode all had scores attached to them. However, we failed to tie subjects' performance on answering questions to their scores, providing them with very little motivation to actually answer questions. In the end, some participants chose to ignore the questions altogether, and it leaves the possibility open that operators ignored the questions because they were focused on the score, not because they were under high workload. However, since all subjects received the same instruction on answering questions, any adverse effects were probably evenly spread between all groups, keeping this a valid measure for workload.

A significant problem with the communication setup manifested itself part-way into the study when several participants believed that their messages to their teammate were not being sent. In some cases this was true: if the "commander" asked them a question, subjects were unable to send messages to their teammate until they gave a proper answer ("proper" means that their answer contained only digits, and they were told this during training). This presented difficulties when, for example, subjects were in the middle of typing a long message and the commander would ask a question. When the message was sent, it would be interpreted as an improper answer to the commander's question, forcing the subject to answer the commander's question and then re-type their message if they so chose. This particular situation was actually somewhat rare, though it seemed to significantly reduce future communication when it did occur.

Another reason subjects believed messages were getting lost was that they were simply missing them among all of the other messages because there was only one chat window. This meant that messages between teammates would often get lost within the distractor messages and commander questions. It was never the intent of this study to artificially make communication more challenging than it had to be, and if given the chance to run the experiment again, we would opt to have a separate chat window to send messages to a teammate.

Due to the way in which the study was advertised and subjects signed up, there was an unintentional bias toward married couples. Out of 16 teams that participated, 9 were married couples. While every team member that participated reported knowing their teammate very well, married couples bring unique dynamics into play that the experiment was not designed to control for. While any performance differences are most likely insignificant, communication might be an area that was impacted more than others.

Perhaps most importantly, travel time for the robots became a significant issue in limiting performance since the maps were fairly large and the robots all started in one location. The speed of the robots could have been increased, but this probably would have adversely affected the cognitive load on the operators due to the greatly increased pace of the experiment. A better approach would have been to more evenly distribute the robots around the maps, giving operators robots that were already close to every building and could be put to work immediately. This would have been a better measure of the capabilities of operators than our experimental setup allowed, but less ecologically valid.

Chapter 5

Conclusions and Future Work

This thesis describes the potential benefits of utilizing operator teams to control multiple remote robots, including a discussion of an experiment in which we empirically demonstrated the benefits of operator teams. This chapter summarizes our conclusions and discusses areas of future research related to our findings.

5.1 Conclusions

Our goal has been to demonstrate that two operators working together as a team can be more effective than two operators working separately. We conducted an experiment to find out if we would be able to see synergistic gains for operator teams controlling multiple robots that would outweigh the costs of coordination among team members.

We found that for a single task controlling homogeneous robots there was approximately linear improvement in performance, as predicted by Steiner's framework. Workload for team members stayed at the same level as workload for operators acting alone.

For operators performing two tasks using heterogeneous robots, effectiveness on the two primary tasks grew approximately linearly. Individual operators saw a significant increase in workload compared to performing one task (an increase of between 25% and 50%), but those working in teams showed no increase in workload over individuals performing one task.

While the constraints of the experiment (task saturation, number of robots, etc.) prevented teams from showing greater than linear improvement in primary task performance, the lower workload while performing two tasks indicates spare capacity that could be utilized to improve primary task performance within a proper environment.

Clearly, not every situation justifies operator teams. However, our results show that if the remote robots can perform sub-tasks independently of one another and the interface supports good situation awareness then there was little impact on primary task performance. This suggests that the potential payoff for using operator teams can be significant, and the impact of coordination on team performance can be low.

Based on the coordination that took place among team members in spite of only a handful of messages sent between team members, it appears that an interface that provides sufficient shared awareness between members of a team can reduce the need for explicit communication; though more targeted experimentation is required to show this empirically. This may be due to the fact that both team members could see what the other team member was doing, and therefore did not need to spend time communicating and getting as many status updates from each other. Previous work has shown that reducing explicit communication can reduce workload and therefore increase performance [54]. However, we are unable to conclude for certain whether the low level of explicit communication was due to effective team behavior or to the coordination afforded by the interface. Another possibility is that the small team size made explicit communication largely unnecessary.

The results are consistent with prior work suggesting that increasing fan-out should not be the only concern when designing human-robot systems (for example, see the arguments posed by Hancock et al. [44]). Workload increases as fan-out increases, which reduces task performance [20]. In our experiment, operators working as a team had lower individual fan-out but similar performance levels than operators working alone. Human-robot system designers should account for this relationship in the systems they build.

Most important, our experiment demonstrates that two people working as a team can gain measurable benefits over two people working as individuals. On average, individuals in teams experienced a greater than 40% reduction in workload while maintaining similar levels of performance on primary tasks.

5.2 Future Work

After analyzing the results from the experiment, there are many questions remaining that we are not able to address using our data.

An important area of future work would be to test if more evenly distributing robots around the map would show even more benefit to operator teams. We feel that the conditions of the experiment artificially limited operator performance due to significant travel time required for the robots, especially the mine-sweepers, resulting in task saturation.

In our experiment, operators in the bombs-and-mines condition were free to switch back and forth between specialization roles, looking for both types of explosives as they pleased. We theorize that this was a benefit to team coordination and performance, as there was less pressure on one individual to perform the entire task (e.g., find all of the bombs or mines themselves), and they could focus on a smaller portion of the task (such as finding all of the mines around a particular building). A question remains as to whether or not operator workload would have been reduced as much if tasks had been strictly assigned, meaning that operators who were searching for bombs could not control mine-sweeper robots and vice-versa.

We also concluded that the reduction in fan-out measurements for individuals on teams contributed to their reduction in workload. Future studies might examine whether or not the reduction in fan-out was due to operators feeling less pressured to complete the task on their own, or if the task could be completed easier if more robots were used. If more robots were available, would operators on teams still push themselves far enough that their workload would increase again? This questions might be answered by providing significantly more robots than can be controlled by both operators, and then looking at how this affects fan-out for individuals on teams.

Future studies might examine whether the reduction in fan-out for team members was due to a lower robot-to-operator ratio, or if it was a self-imposed limitation manifesting as a result of team dynamics. If more robots were available to teams (so that the same robot-to-operator ratio was maintained), would operators increase their workload by trying to control more robots?

References

[1] J. Accot and S. Zhai. Performance evaluation of input devices in trajectory-based tasks: an application of the steering law. In *Proceedings of the SIGCHI conference on human factors in computing systems: The CHI is the limit*, pages 466–472, 1999.

[2] S. Adams. Dilbert, Debuted 1989.

[3] L. Bainbridge. Ironies of automation. *Automatica*, 19:775–780, 1983.

[4] A. Baker and S. Chan. After pilot dies, jet lands safely in newark. *The New York Times*, June 18 2009.

[5] T. Balch and L. E. Parker, editors. *Robot Teams: From diversity to polymorphism*. A. K. Peters Ltd, 2002.

[6] M. C. Berenbaum. What is synergy? *Pharmacological Reviews*, 41(3):422, 1990.

[7] D. J. Bruemmer, D. I. Gertman, C. W. Nielsen, D. A. Few, and W. D. Smart. Supporting complex robot behaviors with simple interaction tools. In N. Sarkar, editor, *Human Robot Interaction*. Advanced Robotics Systems International, Vienna, Austria, 2007.

[8] D. J. Bruemmer, C. W. Nielsen, and D. I. Gertman. How training and experience affect the benefits of autonomy in a dirty-bomb experiment. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 161–168, Amsterdam, The Netherlands, 2008. ACM.

[9] J. L. Burke, R. R. Murphy, M. D. Coovert, and D. L. Riddle. Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human Computer Interaction*, 19(1):85–116, 2004.

[10] L. A. Carroll. Desperately seeking sa. *TAG Attack (TAG SP 127-1)*, 32(3):5–6, 1992.

[11] R. Cassinis, G. Bianco, A. Cavagnini, and P. Ransenigo. Strategies for navigation of robot swarms to be used in landmines detection. In *Proceedings of the Third European Workshop on Advanced Mobile Robots*, 1999.

[12] N. J. Cepeda, A. F. Kramer, and J. Gonzalez de Sather. Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, 37(5):715–730, 2001.

[13] R. Chadwick, D. J. Gillan, and S. L. Pazuchanics. What the robot's camera tells the operator's brain. In N. J. Cooke, H. Pringle, H. Pedersen, and O. Connor, editors, *Human Factors of Remotely Operated Vehicles*, Advances in Human Performance and Cognitive Engineering Research, chapter 26, pages 373–384. Elsevier, 2006.

[14] J. Y. C. Chen, E. C. Haas, and M. J. Barnes. Human performance issues and user interface design for teleoperated robots. *Systems, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37:1231–1245, 2007.

[15] H. A. Colle and R. G. B. Estimating a mental workload redline in a simulated air-to-ground combat mission. *The International Journal of Aviation Psychology*, 15(4):303–319, 2005.

[16] B. E. Collins and H. S. Guetzkow. *A Social Psychology of Group Processes for Decision-Making*. Wiley, 1964.

[17] P. A. Corning. The synergism hypothesis: On the concept of synergy and its role in the evolution of complex systems. *Journal of Social and Evolutionary Systems*, 21(2):133, 1998.

[18] J. W. Crandall and M. A. Goodrich. Experiments in adjustable autonomy. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 3, pages 1624–1629, October 2001.

[19] M. L. Cummings and S. Guerlain. Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1):1–15, February 2007.

[20] M. L. Cummings and C. E. Nehme. Modeling the impact of workload in network centric supervisory control settings. In *2nd Annual Sustaining Performance Under Stress Symposium*, Feb 2009.

[21] M. L. Cummings, C. E. Nehme, J. W. Crandall, and P. Mitchell. Predicting operator capacity for supervisory control of multiple UAVs. *Studies in Computational Intelligence (SCI)*, 70:11–37, 2007.

[22] A. R. Dennis and M. L. Williams. Electronic brainstorming: Theory, research, and future directions. In P. B. Paulus and B. A. Nijstad, editors, *Group Creativity: Innovation Through Collaboration*. Oxford University Press, 2003.

[23] C. Dominguez. Can SA be defined? Technical report, Armstrong Laboratory, 1994.

[24] G. Dudek, J. M., and M. E. A taxonomy of multirobot systems. In T. Balch and L. E. Parker, editors, *Robot teams: From diversity to polymorphism*. A. K. Peters, Ltd, 2002.

[25] M. Endsley. A methodology for the objective measure of situation awareness. In *Situational awareness in aerospece operations (AGARD-CP-478)*, Neuilly-Sur-Seine, France: NATO, 1990. Advisory Group for Aerospace Research and Development (North Atlantic Treaty Organisation).

[26] M. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

[27] M. R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 97–101, 1988.

[28] M. R. Endsley. Automation and situation awareness. In R. Parasuraman and M. Mouloua, editors, *Automation and Situation Awareness*, pages 162–182. Lawrence Erlbaum Associates, 1996.

[29] M. R. Endsley. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland, editors, *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, 2000.

[30] J. A. Espinosa, F. J. Lerch, and R. E. Kraut. Explicit versus implicit coordination mechanisms and task dependencies: One size does not fit all. In E. Salas and S. M. Fiore, editors, *Team Cognition: Understanding the Factors that Drive Process and Performance*. American Psychological Association, 2004.

[31] T. Felin and T. Knudsen. A theory of nascent organizing: On the emergence of collective beliefs and strategy. Marriot School, BYU Working Paper, 2008.

[32] K. Fisher and M. D. Fisher. *The distributed mind: Acheiving high performance through the collective intelligence of knowledge work teams*. AMACOM, 1998.

[33] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381, 1954.

[34] J. M. Flach. Situation awareness: Proceed with caution. *Human Factors*, 37(1):149–157, 1995.

[35] T. W. Fong, M. Allan, X. Bouyssounouse, M. G. Bualat, M. Deans, L. Edwards, L. Fluckiger, L. Keely, S. Y. Lee, and D. Lees. Robotic site survey at Haughton Crater. In *9th International Symposium on Artificial Intelligence, Robotics, and Automation in Space*, February 2008.

[36] B. Gerkey and R. Vaughan. The player project. http://playerstage.sourceforge.net/, Accessed Oct 2009.

[37] K. Goldberg, B. Chen, R. Solomon, S. Bui, B. Farzin, J. Heitler, D. Poon, and G. Smith. Collaborative teleoperation via the internet. In B. Chen, editor, *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, volume 2, pages 2019–2024 vol.2, 2000.

[38] M. A. Goodrich and D. R. Olsen. Seven principles of efficient human robot interaction. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3942–3948 vol.4, 2003.

[39] M. A. Goodrich, M. Quigley, and K. Cosenzo. Task switching and multi-robot teams. In L. E. Parker, F. E. Schneider, and A. C. Shultz, editors, *Proceedings from the 2003 International Workshop on Multi-Robot Systems*, volume 3, pages 185–196. Springer Netherlands, 2005.

[40] W. L. Goodson, R. V. Walters, and R. L. Stites. An artificial neral network for the estimation of tactical situation awareness. Tech. Report USAFSAM-TP-90-9, Brooks Air Force Base, TX: Human Systems Division, 1990.

[41] M. Goold and A. Campbell. Desperately seeking synergy. *Harvard business Review*, 76(5):131–143, Sep 1998.

[42] D. Gopher and E. Donchin. Workload - an examination of the concept. In *Handbook of perception and human performance*, volume 2. Wiley, 1986.

[43] T. L. Griffith and M. Fichman. Social loafing and social facilitation: An empirical test of the cognitive-motivational model of performance. *Basic and Applied Social Psychology*, 10(3):253–271, 1989.

[44] P. A. Hancock, M. Mouloua, R. Gilson, J. Szalma, and T. Oron-Gilad. Is the uav control ratio the right question? *Ergonomics in Design*, 15(1):7–30, 2007.

[45] S. Hart and L. Staveland. Development of NASA-TLX (Task load index): Results of empirical and theoreticalresearch. In P. Hancock and N. Meshkati, editors, *Human mental workload*, pages 139–183. Amsterdam: North Holland Press, 1988.

[46] D. O. Hebb. Drives and the cns (conceptual nervous system). *Psychological Review*, 62:243–254, 1955.

[47] R. L. Helmreich, A. C. Merrit, and J. A. Wilhelm. The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology*, 9(1):19–32, 1999.

[48] D. Kahneman. *Attention and Effort*. Prentice Hall, 1973.

[49] B. Latane, K. Williams, and S. Harkins. Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6):822–832, 1979.

[50] J. D. Lee and K. A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004.

[51] K. Lerman, A. Galstyan, A. Martinoli, and A. Ijspeert. A macroscopic analytical model of collaboration in distributed robotic systems. *Artificial Life*, 7(4):375–393, 2002.

[52] M. Lewis, J. Wang, and P. Scerri. Teamwork coordination for realistically complex multirobot systems. In *NATO Symposium on Human Factors of Uninhabited Military Vehicles as Force Multipliers*, 2006.

[53] G. D. Logan. The time it takes to switch attention. *Psychometric Bulletin and Review*, 12(4):647–653, 2005.

[54] J. Macmillan, E. E. Entine, and D. Serfaty. Communication overhead: The hidden cost of team cognition. In E. Salas and S. M. Fiore, editors, *Team Cognition*. American Psychological Association, 2004.

[55] L. K. Michaelson, W. E. Watson, and R. H. Black. A realistic test of individual versus group consensus decision making. *Journal of Applied Psychology*, 74(5):834–839, 1989.

[56] L. K. Michaelson, W. E. Watson, A. Schwartzkopf, and R. H. Black. Group decision making: How you frame the question determines what you find. *Journal of Applied Psychology*, 77(1):106–108, 1992.

[57] R. R. Murphy. Humans, robots, rubble, and research. *Interactions*, 12:37–39, 2005.

[58] R. R. Murphy and J. L. Burke. Up from the rubble: Lessons learned about HRI from search and rescue. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, pages 437–441, 2005.

[59] J. V. Nickerson and S. S. Skiena. Attention and communication: Decision scenarios for teleoperating robots. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, page 295, 2005.

[60] C. W. Nielsen and D. J. Bruemmer. Hiding the system from the user: Moving from complex mental models to elegant metaphors. In *The 16th IEEE International symposium on Robot and human interactive communication*, pages 756–761, 2007.

[61] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *Robotics, IEEE Transactions on*, 23(5):927–941, 2007.

[62] B. A. Nijstad, W. Stroebe, and H. F. M. Lodewijkx. Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, 39(6):531–548, 2003.

[63] D. R. Olsen and S. B. Wood. Fan-out: Measuring human control of multiple robots. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 231–238, Vienna, Austria, 2004. ACM.

[64] J. M. Orasanu. Decision making in the cockpit. In E. L. Wiener, R. L. Helmreich, and B. G. Kanki, editors, *Cockpit Resource Management*. Elsevier Science & Technology Books, 1995.

[65] R. Parasuraman, S. Galster, P. Squire, H. Furukawa, and C. Miller. A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with Roboflag. *IEEE transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 35:481–492, 2005.

[66] L. A. Penner and J. P. Craiger. The weakest link: The performance of individual team members. In R. W. Swezey and E. Salas, editors, *Teams: Their training and performance*. Ablex Publishing Corporation, 1992.

[67] A. Pinsonneault, H. Barki, R. B. Gallupe, and N. Hoppen. Electronic brainstorming: The illusion of productivity. *Information Systems Research*, 10(2):110–133, 1999.

[68] K. M. Propp. In search of the assembly bonus effect. *Human Communication Research*, 29(4):600–606, 2006.

[69] M. J. Rosenberg. The conditions and consequences of evaluation apprehension. In R. Rosenthal and R. L. Rosnow, editors, *Artifacts in Behavioral Reserach*. Academic Press, 1969.

[70] N. B. Sarter and D. D. Woods. Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, 1(1):45–57, 1991.

[71] N. B. Sarter, D. D. Woods, and C. E. Billings. Automation surprises. In *Handbook of Human Factors & Ergonomics*. Wiley, 2nd edition, 1997.

[72] D. Serfaty and E. E. Entine. Team adaptation and co-ordination training. In R. Flin, M. Salas, E. an Strub, and L. Martin, editors, *Decision Making Under Stress: Emerging Themes and Applications*. Ashgate Publishing, 1997.

[73] T. B. Sheridan. *Telerobotics, automation, and human supervisory control*. MIT Press, 1992.

[74] J. M. Smith. Group selection and kin selection. *Nature*, 201(1):1145–1147, 1964.

[75] I. D. Steiner. *Group Process and Productivity*. Academic Press, New York, NY, 1972.

[76] R. S. Tindale and J. R. Larson. Assembly bonus effect or typical group performance? A comment on Michaelsen, Watson, and Black (1989). *Journal of Applied Psychology*, 77(1):102–105, 1992.

[77] J. A. Wagner III. Studies of individualism-collectivism: Effects on cooperation in groups. *The Acadamy of Management Journal*, 38(1):152–172, 1995.

[78] C. D. Wickens and J. G. Hollands. *Engineering psychology and human performance*. Prentice Hall, 3rd edition, 2000.

[79] C. D. Wickens, A. S. Mavor, and J. P. McGee, editors. *Flight to the future: Human factors of air traffic control*. National Academy Press, 1997.

[80] E. Wiener. Cockpit automation. In E. Wiener and D. C. Nagel, editors, *Human Factors in Aviation*. Academic Press, 1989.

[81] D. D. Woods, J. Tittle, M. Feil, and A. Roesler. Envisioning human-robot coordination in future operations. In *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, volume 34, 2004.

[82] H. A. Yanco and J. L. Drury. Classifying human-robot interaction: An updated taxonomy. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2841–2846, 2004.

[83] H. A. Yanco and J. L. Drury. “Where am I?” Acquiring situation awareness using a remote robot platform. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 2835–2840, 2004.

[84] H. A. Yanco, J. L. Drury, and J. Scholtz. Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human-Computer Interaction*, 19(1):117–149, June 2004.

[85] S. J. Zaccaro, J. Gualtieri, and D. Minionis. Task cohesion as a facilitator of team decision making under temporal urgency. *Military Psychology*, 7(2):77–94, 1995.

[86] R. B. Zajonc. Social facilitation. *Science*, 149(3681):269–274, 1965.

Appendix A

Design and Implementation of the Robot Simulator and Interface

In order to run an experiment in which multiple operators could collaboratively control multiple robots as a team, we first needed a software system capable of simulating a large number of robots. Any robot simulator needed to be able to dynamically update world maps and share them across more than a dozen simulated robots. We also wanted the simulator to have some basic robot autonomy available, including moving between waypoints and planning new paths. In addition to all of this, the simulator needed to be fast enough to run on modest hardware while still providing two operators with real-time updates.

Commercial options were quickly discarded due to concerns about complexity and lack of customization capability. The Player Project [36] was seriously considered because it is an open-source implementation of a robot simulator that includes several built-in automated behaviors. After spending considerable time and effort attempting to make the needed adjustments to make the simulator meet our needs, we determined that the basic design of the software did not fit with our needs. Faced with compromising on the scope of our experiment or significant customization difficulties, we decided it would be easier to design and build a new simulator from scratch that was customized to our needs.

A.1 Overview

In order to meet our particular needs, we designed and developed an entirely new 2D robotic simulator that uses simple physics to represent robots and the environment around them. We used this simulator to run a low-fidelity simulation of a large number of robots in a shared environment (robots were simply treated as rectangles, greatly reducing the computational requirements). Running a low-fidelity simulation makes it possible to simulate dozens of robots in a shared environment while still running on modest hardware.

The interface we used to connect to the simulator is based on an interface developed primarily at the Idaho National Lab (INL). The INL's interface displays a robot in a 3D virtual environment that allows for better situation awareness than the more typical camera-based interfaces [61]. We expanded the capabilities of the interface to allow for command and control of an arbitrary number of robots, allow multiple interfaces to share the same virtual environment through a network connection, and added numerous other minor features to fit our needs.

Both the simulator and the interface were written in C# using OpenGL to render most of the graphics. We implemented a client-server architecture to allow multiple instances of the interface to be run on multiple machines and still be synchronized with each other. The simulator acts as the server, pushing out state information about the robots while receiving and executing commands from the client interfaces. A basic overview is shown in Figure A.1.

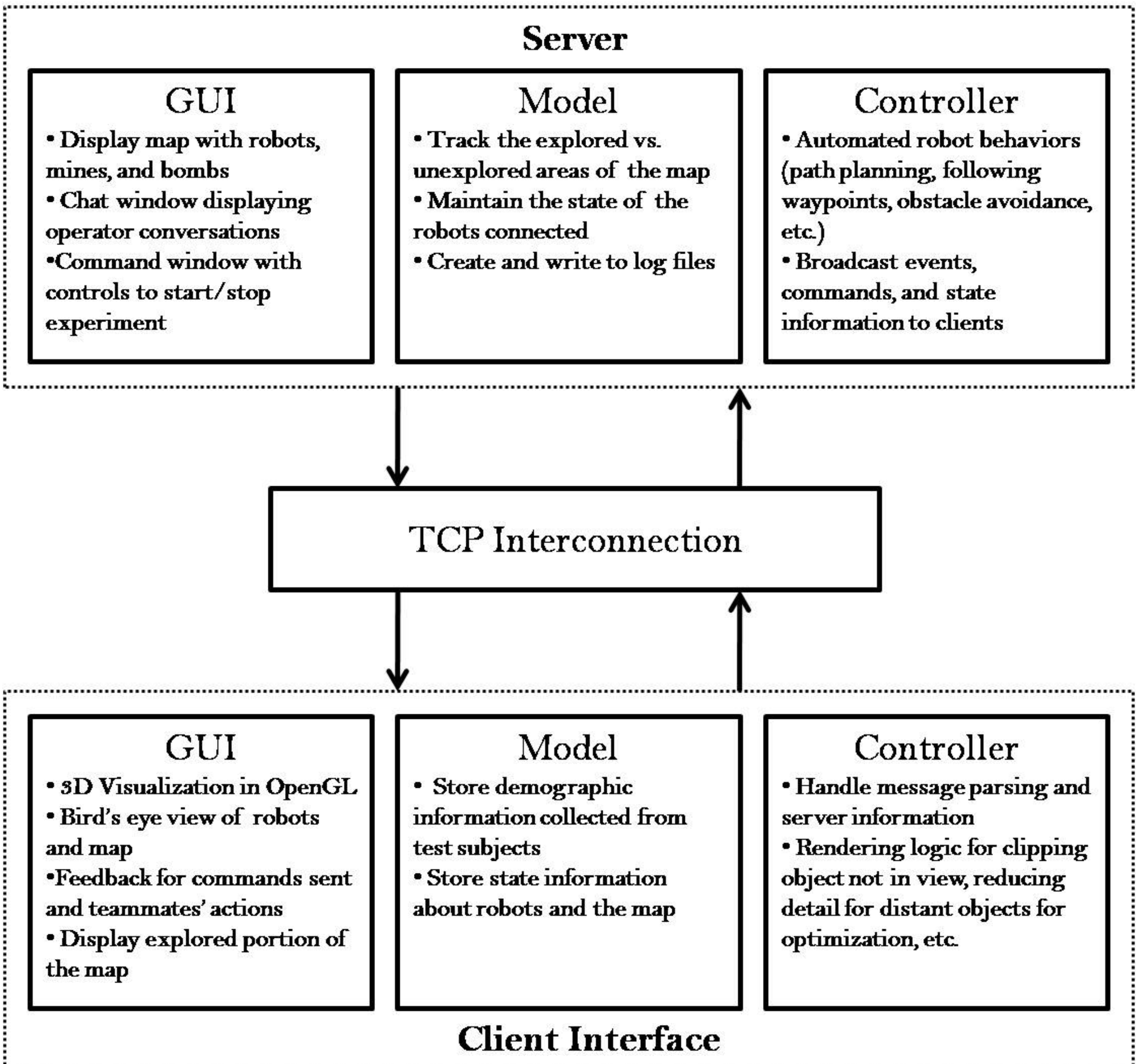

Figure A.1: Basic system outline for interface and simulator.

A.2 Implementation

The client interface contains approximately 27,000 lines of code (LOC), and the simulator has nearly 10,000 LOC.

The system has a built-in chat capability where each interface has a chat window that receives individualized messages generated by the server as well as messages from other client interfaces (routed through the server). This feature was used in our experiment to allow communication between team members as well as provide a secondary task (answering questions via the chat window).

The server/client connection is made using TCP, with an initial auto-discovery enabled by a UDP broadcast from the server. The client interfaces never connect directly to each other, meaning all data goes through the server (even though some messages are directed at another interface and not the server). Bandwidth requirements are fairly low, with most of the data consisting of either coordinates or ASCII plain-text messages.

Another important feature to our experiment is the logging feature. Every action, command, and status update is written to one of several log files. For our experiment, these logs were used to obtain measurements using a Python script. The script parses through the logs, extracts data, creates charts, and calculates significance and correlation values.

The simulator implements some basic automation for the robots, including path planning, collision avoidance, and waypoint following. Our simulator plans paths using a simple A* algorithm that searches through nodes generated by the probabilistic roadmaps method (nodes are placed randomly on the map within simple constraints). The robots replan paths every few seconds so that they can dynamically respond to newly discovered obstacles (obstacles are revealed in front of the robots as they move around the map). Collision avoidance and waypoint following are achieved by using potential fields to either attract or repel robots.

Appendix B

Subjective Workload Assessment Survey

Survey Given to Measure Subjective Workload

1) How much mental activity was required? Was the task easy or demanding?

1 Easy	4 Somewhat Demanding
2 Somewhat Easy	5 Very Demanding
3 Undecided	

2) How much time pressure did you feel due to the rate or pace at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

1 Very Slow/Leisurely	4 Somewhat Fast-paced
2 Somewhat Slow/Leisurely	5 Very Fast-paced
3 Undecided	

3) How hard did you have to work mentally to accomplish your level of performance?

1 Not very hard
2
3 Somewhat hard
4
5 Very Hard

4) How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

1 Unsuccessful	4 Somewhat Successful
2 Somewhat Unsuccessful	5 Very Successful
3 Satisfactory	

5) How satisfied were you with your performance in accomplishing the goals of the task?

1 Unsatisfied
2
3 Somewhat Satisfied
4
5 Very Satisfied

6) How irritated and stressed versus content and relaxed did you feel during the task?

1 Irritated/Stressed most of the time
2
3
4
5 Content/Relaxed most of the time