Situation Awareness and Task Performance in Robot-Assisted Technical Search: Bujold Goes to Bridgeport

Jennifer L. Burke, Professor Robin R. Murphy
Center for Robot-Assisted Search and Rescue
University of South Florida

Keywords: human-robot interaction, search and rescue, professional service robots, mobile robots, artificial intelligence, situation awareness, multi-operator single robot, teleoperation

Abstract

This article reports on a field study conducted in 2002 in Bridgeport, Connecticut, where 28 multi-operator single robot (MOSR) teams were videotaped as they teleoperated a rescue robot through an apartment in a collapsed building in search of a victim mannequin. Team communication analysis was conducted using the Robot-Assisted Search and Rescue Coding Scheme (RASAR-CS), in conjunction with the administration of three instruments measuring situation awareness and task performance. The major findings were that 1) rescuer teams with high situation awareness operators were 9 times more likely to find victims than rescue teams with low situation awareness operators, and scored 26% higher on ratings of task performance; 2) operators spent 63% of their time on perceiving and comprehending (SA Level 1 and 2 activities), and only 28% on planning, projecting and problem-solving (SA Level 3 activities); 3) the shared mental model held by the robot operator and tether manager consisted of at least three distinct types of knowledge: *the environment*, *the robot's situatedness*, and *search strategy*. The study validates a computational model of SA formation in MOSR teams, which is expected to serve as a foundation for artificial intelligence methods in awareness. The results also suggest that major advances are needed in sensors and sensor interpretation to facilitate lower level SA activities so that the operator will more rapidly have a higher level SA; and that ways (e.g., training, software agents) must be found to facilitate appropriate communication to support productive team processes.

I. INTRODUCTION

Robot-assisted search and rescue was identified by the 2002 DARPA/NSF Study on Human-Robot Interaction [1] as an exemplar domain for studying HRI. It is a real-world application as seen by the first use of rescue robots at the World Trade Center disaster [2] . In urban search and rescue, robots the size of shoeboxes or carry-on luggage are inserted into voids in the rubble pile of a collapsed building, subway, or other urban structure to conduct the various missions. Humans teleoperate the robots 5 to 30 meters into the highly confined interior of the void, looking for signs of survivors (e.g., the technical search mission), the state of the collapse in terms of victim extrication (e.g., the rescue mission), and whether it is safe for rescuers to enter (e.g., the hazardous materials mission). If victims are found, rescuers can triage the victim, send the robot down with lighting, food, and tubing transporting water or air to the victim, and maintain conversations with the victim (e.g., the medical mission). The teleoperators work outdoors with little protection from weather conditions, rarely sleep for the first 52 hours, and are under significant emotional pressure to find survivors [2]. The demanding, unpredictable nature of search and rescue is similar to the domains of military operations in urban terrains, SWAT and bomb-disposal, and battlefield medicine, and is harder than planetary exploration[3, 4]. Therefore results from robot-assisted search are expected to be widely applicable to any human-robot endeavor for safety or security activities.

While the robot missions are varied, the levels and types of interactions between the robot and humans are even more diverse. “Behind” the robot is a hierarchy of trained rescue personnel working with robots and converting that data into information to be passed up to the team manager, the task force manager, and finally to the incident commander. Rescue workers are likely to have a spectrum of training, knowledge, expectations, and comfort with robots. “In front” of the robot is a trapped and frightened victim who is unlikely to have any knowledge about robots. The rescue robots are currently teleoperated, but semi-autonomy and mixed-initiative control is being introduced, allowing for studies to both continue to follow the impact of progress in artificial intelligence, and to influence the technology roadmap.

The scope of this article is on the human-robot interaction in the first rung in the information hierarchy: the operators directly interacting with the robots. This rung is particularly interesting because since 2001, rescue robots are a multi-operator single robot (MOSR) domain. At least two operators are used per robot, i.e., a 2:1 human to robot ratio. The 2:1 ratio is the result of logistics, accepted practice, and three prior studies in HRI for search and rescue. Since most search and rescue robots are connected to a control station via a tether that also serves as a safety line, a human generally is required to act as the *tether manager*, lowering the robot and keeping the line untangled while the other human sits at the operator control unit and serves as the *robot operator*. The use of two operators per robot is not a manpower issue since search teams naturally work in pairs (the “buddy system”) for safety. While it might be possible to improve the robot hardware so as to combine the tether manager and robot operator roles, three studies with emergency responders in field conditions, one in July, 2001 with Florida Task Force

3[5], one an analysis of HRI at the World Trade Center deployment [2], and the third a study in Miami in December, 2001[6], suggest that this might be a false economy. The three studies showed that two operators working together to interpret the video data did much better than a single operator, who often missed victims or remains. The Miami study was particularly interesting because it showed that the two operators were talking with each other to cognitively build a *shared mental model* of the search environment and the overall situation (e.g., *situation awareness)*. Therefore, simply eliminating the physical need for the second person might actually reduce overall performance.

The approach taken to human-robot interaction in this article is motivated by the central tenet of our research: that the study of situation awareness (SA) is not only essential to moving from multi-operator single robot to single operator single robot control, but also for moving from teleoperation to autonomy and for human interaction with autonomous robots in general. First, in teleoperation and semi-autonomous control of mobile robots, the human is the agent that builds and maintains SA for the system. Unless SA is understood, the appropriate divisions of responsibilities between human and robots, useful user interfaces, and effective training procedures are unlikely to be created. Next, semi- and fully autonomous robots will need to be able to construct and maintain their own SA. This, in turn, means that the SA required for a task must be explicitly extracted and represented so it can be programmed into a robot. Finally, as more humans become partners with robots, the information technology system at large will need to assist the humans, who may not be robot experts, in creating SA. To facilitate the formation of effective team processes and to make robots more intelligent, situation awareness must be understood. Whatever the effective teams are doing needs to be captured in a way that lends itself for encapsulation in an autonomous or semi-autonomous system, and what they appear to be doing is constructing situation awareness.

Following our research objective, this article details a field study conducted in 2002 with emergency responders in Bridgeport, Connecticut that was designed to study the relationship of situation awareness to task performance and to identify the components and process of constructing situation awareness. The previous field studies cited above were more ethnographic in nature: researchers observed emergency responders using the robots whenever they found a suitable void, training of the personnel using the robots varied widely from practically no training to experts, and data was collected from different operators in different voids where the ground truth was not always known. These studies provided only a coarse understanding of what led to good task performance. In this study, 175 search and rescue professionals were trained using a 2:1 team protocol on how to use the robots through a formal one hour awareness class prior to a 24 hour training exercise in a collapsed building. During the actual exercise, the robots were "on call" and used as needed but also at a "station," where one robot and a specific area of the collapsed building were set aside for data collection, providing a repeatable task scenario and ground truth as to the status and location of the robot at any given time. As responders went on breaks or were not needed for activities during the exercise, they rotated through the robot station. Twenty-eight of the 175 responders were able to participate in the robot station throughout the exercise, particularly as the pace of regular operations slowed down in the early morning hours. This is the first known controlled study of human-robot interaction with actual robots and end users operating in realistic conditions for an extended period of time.

The article is organized as follows. It begins with a review in Section II of the related work in human-robot interaction that has investigated the relationship of situation awareness to task performance, team communication, and shared mental models. Section III presents 1) the model of SA formation, where the two operators have individual mental models appropriate to their role but also create a common, dynamic model of the overall situation by talking with each other, and 2) the five hypotheses that motivated the study. The Connecticut study was far more controlled than the previous field studies and yielded much more detailed data, and the methodology used to collect, code, and analyze the task performance and verbal communication data is summarized in Section IV. The results and findings are reported in Section V. Section VI discusses the validation of the model of SA formation and contents of the shared mental model, the ramifications for training and cognitive augmentation, and limitations of the study. The article closes with the conclusions presented in Section VII.

II. RELATED WORK

Studies in human-robot interaction are often categorized not by cognitive functionality but rather according to the domain functionality of the robots in question: industrial, professional service, or personal service robots [7]. USAR robots fall into the professional service category, along with those in the medical field, the military, and space applications (e.g., [8-10]), where robots are intended to work with a human to meet the human's goals. SA has a long research history as a key component of human-machine systems in military and civilian aviation and space exploration, with research typically concentrating on the relationship between SA and task performance or on the psychological constructs underlying human-machine performance—situation awareness, shared mental models, team processes, and metrics and methodologies. This article follows human-machine research convention and organizes HRI studies by cognitive functionality. To summarize, there have been only six studies identified to date which attempt to link SA and task performance in HRI; one study was a simulation exploring telepresence [11] with little relevance to this effort, two studies were derived from the RoboCup Rescue Competition with robots and humans operating in a grossly simplified USAR setting [12, 13], and three were field studies conducted by the Center for Robot-Assisted Search and Rescue (CRASAR) [2, 5, 6]. CRASAR field studies provide the foundation for this article. In addition, team communication work in [14] reinforces the central tenets of this article. Studies concentrating on improving task performance in human-robot teams appear to be universally robot-centric, testing either the robot's abilities or some component of the robot as a factor influencing performance, and ignore the role of the human. This provides no insight into the larger human-robot team. Professional service robots assist people in attaining their professional goals; therefore, a logical metric of human-robot performance is whether the person's goal is achieved. However, the question of how to measure human-robot interaction is largely ignored, with the notable exceptions of [10, 15-18]. Since these methods are generally usability- or evaluation-focused, this article relies on methodologies taken from the psychology community and encapsulated in the RASAR-CS scheme described in Section IV.

A. Situation Awareness Linked with Task Performance

Only six studies have been found that examine situation awareness in conjunction with task performance. Riley, Kaber and Draper [11] looked at SA, performance and attention allocation in a study investigating the relationship between telepresence and teleoperation performance in a simulated mine disposal task. Their study, a Wizard of Oz experiment (i.e., one in which an apparently autonomous robot is controlled by the experimenter) in a laboratory setting, examined SA as a possible indicator of telepresence. In contrast, this article explores SA as a predictor of task performance.

The annual Robocup Rescue Competition has been used to compare human-robot teams' performance in a rescue-oriented domain [12, 13]. However, the physical setting and conditions are quite different from those experienced at a disaster site, and the robots used are not fieldable, i.e., they are designed for short competition rounds in the NIST testbed rather than for a true disaster environment [19]. Moreover, the people on these teams are robot developers rather than rescue professionals, and have neither the training nor the skills of the intended end-users.

Field studies conducted with true end-users in ecologically valid disaster response settings offer a more realistic look at human-robot performance in terms of current capabilities. Three field studies conducted by the Center for Robot-Assisted Search and Rescue (CRASAR) prior to this study recorded real robot-user interaction as it occurred between team members and a single robot to inform the development of coordinated human-robot systems within the organizational structure of USAR [2, 5, 6]. In each of these, situation awareness [20] is a key construct for understanding (and improving) human-robot interaction.

The first CRASAR field study was an ethnographic study of Florida Task Force 3 members using robots to search for a victim [5], while the second was an analysis of data collected during the use of rescue robots at the World Trade Center disaster. The Florida Task Force 3 study suggested that two operators are needed to interpret multiple sensor data while navigating due to the simultaneous nature of activities described as part of the technical search task (searching for victims and structural inspection). Casper and Murphy's [2] analysis of video data collected during the World Trade Center disaster response found that operators' lack of awareness regarding the state and situatedness of the robot in the rubble impacted performance of human-robot teams. Operators also had difficulty linking current information obtained from the robot to existing knowledge or experience. Both the Florida Task Force and World Trade Center human-robot interaction studies reveal difficulties in operator teleproprioception and telekinesthesis, consistent with the problems described in [21].

The third study [6] identified situation awareness and team communication as critical elements in human-robot interaction. That study was conducted in Miami during a 16-hour high-fidelity USAR disaster response drill where five operators were observed and video-recorded as they teleoperated robots in technical search operations. Operator situation awareness and technical search team interaction were examined using the Robot-Assisted Search and Rescue

Coding Scheme (RASAR-CS), a systematic coding scheme designed for this research. The findings indicated that operators spent significantly more time gathering information about the state of the robot and the state of the environment than they did navigating the robot. Operators had difficulty integrating the robot's view into their understanding of the search and rescue site. They compensated for this lack of situation awareness by communicating with tether managers and other team members at the site, attempting to gather information that would provide a more complete mental model of the site. They also worked with these team members to develop search strategies. Indeed, operators rated as having high situation awareness talked more with team members than operators rated as having low situation awareness, especially about search strategies and information gathered from synthesizing what was seen through the robot's eye view with what was already known about the environment.

.

B. Team Communication and Shared Mental Models

Jones and Hinds' qualitative analysis [22] of police SWAT teams (an "extreme team" domain similar to USAR) is the closest work conceptually to the goals of this article. Jones and Hinds explored the importance of team communication in the development of a shared mental model (which they termed "common ground"), and noted the implications for SWAT team performance. They observed police SWAT teams in training exercises, and identified leader roles in establishing common ground and coordinating distributed team member actions as factors transferable to system design for coordinating distributed robots. Jones and Hinds' work studied distributed SWAT teams (people) to model a team of distributed robots that could work together in similar fashion (but not specifically with police SWAT teams). In contrast, the field study in this article recorded real robot-user interaction as it occurred between team members and a single robot to inform the development of coordinated human-robot systems within the organizational structure of USAR. This article uses the findings regarding the criticality of shared awareness in team-based, dynamic work domains as a justification for exploring team communications in the USAR domain.

C. Task Performance

The research goals of efforts investigating task performance in human-robot systems can be grouped into two categories, neither of which investigates a link between situation awareness and task performance and so are not useful for this article. In the first category are efforts which compare human performance to that of a robot or computer algorithm [17, 18]. These studies focus heavily on the robot's usability rather than on the psychological constructs underlying human-robot task performance. For example, Lumelsky [17] compared operator performance to that of a computer-based motion-planning algorithm in a series of laboratory experiments with participants teleoperating an industrial robot arm. He blamed the participants' poor performance on human limitations in spatial orientation and interpretation of spatial data. The goal of this research was more to justify human-robot task allocation rather than to examine man-machine system performance. Crandall, Nielsen and Goodrich [18] also compared human performance on a simulated goal-finding exploration task to that of a computer algorithm. Their goal was to validate a workload metric (robot attention demand – a combination of neglect tolerance and interface efficiency) as a predictor of robot team performance.

In the second category, efforts examine the effects of various robot characteristics (level of autonomy, interface type, interaction modality) on human-robot system performance [10, 15, 16]. In a usability evaluation of a robot mission-planning wizard for a hybrid deliberative and reactive control system (MissionLab), Endo, MacKenzie and Arkin [10] conducted usability experiments with 29 participants, looking at performance outcomes such as speed, accuracy, and user attitudes regarding ease of use with and without the wizard. Marble, Bruemmer and Few [15] also conducted usability experiments, using both novice and experienced teleoperators to evaluate a mixed-initiative robotic system in a laboratory-based search task. Their focus was on the effects of levels of autonomy on performance and perceived ease of use. Perzanowski et al. [16] conducted a Wizard-of-Oz pilot study with 5 participants interacting with a robot to perform a search task. Their goal was to examine interaction modality style of the participants rather than their performance working with the robot.

D. Metrics and Methodologies

Most of the human-robot interaction studies cited above were conducted in laboratory settings (with [12, 13] being the exception), and so do not contribute to the experimental methodology described in Section IV. While some studies used real robots [12, 13, 15, 17], most used simulations or Wizard-of-Oz techniques in their experiments. Study designs and methodologies are varied. Some studies have used a control group/treatment group design [10], which allows use of quantitative analyses (e.g., ANOVA) to compare differences between groups on metrics such as time-to-completion, accuracy, and ease of use; others apparently have used a within subjects design to compare performance differences on task-based measures (e.g., path length) under different treatment conditions [15, 17]. Sample sizes are small, and study participants range from undergraduate college students [11] to people with technical backgrounds [10]; many studies did not specify participant characteristics or the intended population sample they wished to draw from. To our knowledge, no one is drawing participants from a sample population of targeted end-users, as in this study. In many cases, experimental design is lacking. A few studies have used psychological research methods such as critical events analysis or protocol analysis techniques [12, 13, 16], which involve coding events or user actions. However, the coding is generally on a broad level and is used to inform qualitative analyses of the events or action in question.

Psychological research methods include a wide variety of experimental and nonexperimental designs appropriate for use in both laboratory and field settings. Measures of human behavior, attitude and cognition can be created that are reliable, valid, and generalizable across settings, populations, and research domains. This study draws on psychological research methodologies in an observational field study, and uses communication coding techniques and reliable measures to conduct quantitative analyses of operator communications and performance.

III. APPROACH

This article focuses on establishing the link between SA and task performance, with a secondary interest in determining the components leading to the construction and maintenance of SA. The approach is to posit and validate a cognitively plausible model of how SA is formed in robot-assisted technical search, then use that model to derive specific hypotheses and experiments. This section first describes the model, then generates a set of five hypotheses which are the specific objectives of the study.

A. Model of SA Formation

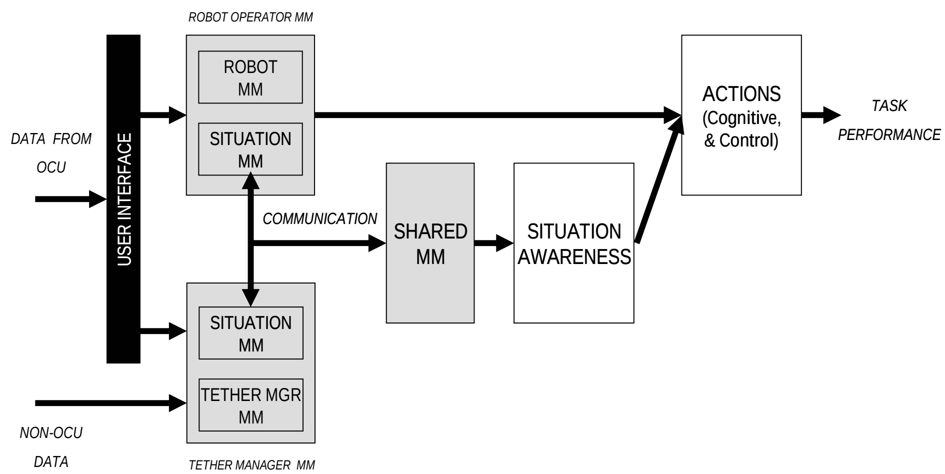

Figure 1 Model of formation of situation awareness (SA) in robot-assisted technical search.

Figure 1 shows the model of SA formation for robot-assisted technical search. It is a synthesis of the results of the Miami study and is consistent with the literature on mental models and situation awareness. The model is explained following the diagram, left to right, below.

The human-robot team consists of two operators and one robot. The operators are the *Robot Operator* and the *Tether Manager*, each of whom has a *mental model (MM)*. The Robot Operator is the person who directly operates the robot. The Tether Manager handles the robot tether or safety line. The input to the robot operator is data from the robot processed through the *Operator Control Unit* (OCU). The OCU serves as the *user interface*. Both the Robot Operator and the Tether Manager can view the OCU; the Robot Operator generally looks constantly at the OCU, and the Tether Manager typically only looks at the OCU intermittently as his responsibilities permit. This means that the Tether Manager has a different perspective on task progress and the overall situation since he can see the exterior (and sometimes what the robot is doing), and can feel the robot's movements through the tether (e.g., whether it is moving forward and drawing more line).

The operators develop two types of individual internal representations called *mental models (MM).* One type is the *role-specific mental model*. Hinsz [23] defines a mental model as "…an individual's representation of a system, and the individual's interaction with the system, with particular focus on how the individual's interactions with the system leads to the outcomes of interest" (p. 202). The role of the Robot Operator requires a *robot mental model* that reflects how he represents the robot technology while the Tether Manager has a *tether manager mental model* that encapsulates how to handle the tether. In addition to the role-specific mental model, each operator has a *situation mental model*. Situation models extend beyond the more static mental models of the system, task and team to represent the dynamic, present state of the system [24]. For example, the Robot Operator has a mental model of how a robot functions, and how his actions affect the robot's actions in performing a task based on training and experience; but he also has a more dynamic (situation) mental model of how the robot is functioning right now in this particular environment.

The model posits a *shared mental model* as a fusion of the two individual situation mental models into a common ground [22]. The shared mental model concept follows Orasanu [25], who suggested that teams faced with novel situations or emergencies (as is often the case with USAR teams) must also develop *shared* situation models for the specific problem. Important parts of these shared situation models according to Orasanu include shared understanding of the problem, goals, information cues, strategies and member roles. The model in this article goes further and assumes that fusion takes place via communication between the human team members, which is consistent with research on team processes and mental models [25-28].

In the model, situation awareness is constructed from the shared mental model and provides the basis for the *actions* of the Robot Operator. This is consistent with Endsley who succinctly stated in [29], "Mental models are the key enablers of Levels 2 and 3 SA" (p.23). The actions of the Robot Operator may be either *cognitive* (e.g., identification of a victim) or *control* (e.g., navigate the robot to a new configuration or location). These actions then lead to *task performance*, which is the discovery of victims for this article.

It should be noted that while the quality of the user interface will have some impact on SA, this study is restricted to the formation of the shared mental model through team communication. This restriction permits the study scope to be tractable. It also ensures that this study will produce results about the human-robot interaction which entails two different roles for humans.

B. Hypotheses

Based on the results of CRASAR field studies, the existing body of literature connecting SA with task performance, and the model posited above, this article theorizes that the relationships between situation awareness, team communications, and shared mental models will positively impact human-robot team performance. Based on the Miami study, this article first hypothesizes that operators will spend more time trying to gain situation awareness than they spend performing tasks requiring situation awareness, and then poses that in fact, they must have this information in order to perform the cognitive tasks of search and navigation. If these hypotheses are confirmed, it indicates that the lack of perception is the major bottleneck in task performance. Based on the model, the article further hypothesizes that goal-related team communication will aid in the development of shared mental models of the situation, resulting in greater operator situation awareness and subsequently better operator task performance. The five formal hypotheses and how they are related are detailed below.

It should be noted that “communication” for the purposes of these hypotheses means verbal communication. At first this may seem unduly restrictive. However, USAR teams often work in environments that require extensive safety gear (e.g., helmets, personal protective equipment) which hides facial expressions and body language; in many cases their only common ground is a shared visual image, and (similar to cockpit crews) the only way to confirm their shared mental models of the task and their roles in that task is to talk to each other. Second, research has shown that conversations among high performing cockpit crews were characterized by great homogeneity [28]; the development of conventionalized speech patterns that facilitate coordination has been linked with high performance because team members interact in predictable ways.

1) Building and Maintaining Shared Mental Models and SA: The first two hypotheses concern building and maintaining situation awareness. SA consists of two kinds of knowledge: knowing what is happening around you, and understanding what it means to you, both now and in the future. The concept is associated with operational situations, where one must have SA for a particular job or function [29]. Following [6], this article uses Endsley's three-level model [20], which defines situation awareness as "…the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (p.97). Perception (Level 1) is detection of sensory information: the perception of elements in the environment within a volume of time and space. Comprehension (Level 2) is divided into two subcategories, identification and interpretation. Identification is defined as comprehension of perceived cues in terms of subjective meaning: e.g., identifying objects, locations and victims. Interpretation is defined as comprehension of perceived cues in terms of objective significance or importance to the current situation. Projection (Level 3) is defined as the projection of future situation events and dynamics through projecting, generating and activating solutions/plans. Jones and Endsley's [30] study of situation awareness in the aviation domain found that more than 90% of pilot errors were associated with Levels 1 and 2 situation awareness. In our previous studies, operators had difficulty building and maintaining the first two levels of SA, and spent more time talking about what they were seeing and what it meant than they did using that information to plan and conduct their search. Therefore, we hypothesize the following:

Hypothesis 1: More operator communications will be related to identification and comprehension (Levels 1 & 2 SA) than to projecting, planning & problem-solving (Level 3 SA).

If this hypothesis is supported, then perception is indeed the bottleneck in building and maintaining SA. Since machine perception is notoriously difficult to automate, this would cast doubt on the practicality of taking the human out-of-the-loop without major advances in sensors and sensing.

As stated in [29], an operator's ability to form projections and plans (Level 3 SA) is only as good as his understanding of the situation, i.e., without Level 2 SA (comprehension), there is no Level 3 SA. Thus, we further hypothesize that:

Hypothesis 2: Operator communications pertaining to Levels 1 (perception) and 2 (comprehension) SA will be related to communications denoting Level 3 SA (planning, projection and problem-solving).

Specifically, this hypothesis predicts that operator communications about the environment and information synthesis (statement categories described in Section IV) will be critical in developing search strategy, and that operator communications about the robot's state and situatedness, in addition to the environment and information synthesis, will be critical in navigation. In other words, talking about perceiving and understanding the environment (SA levels 1 and 2) will be critical not only for the immediate activities of moving about and looking around, but also for making critical decisions regarding the search process. These decisions are embedded in performance [14]. Like cockpit crews or air surveillance teams, USAR teams are performance-oriented, that is, they exist to perform a task. The decisions they make are part of an ongoing larger activity.

2) Team Communication and SA: Hypotheses 3 and 4 motivate the research on team processes and mental models and support *team communication* as integral to the development of shared mental models in teams. In a study of military command and control exercises, [26] found that frequent communications between team members about the work context and situation, work process and domain-specific information were needed to maintain shared situation awareness in dynamic, constraint-bound contexts. In a study of the cognitive functions of cockpit crew member communication, [25] found that captains of high performing crews explicitly stated more plans, provided more explanations, and made more predictions, which were articulated for the whole crew. This enabled crew members to contribute relevant information or strategies from their specialized perspectives, and to interpret requests and commands unambiguously. [27] studied the effects of fatigue on crew coordination and performance, and suggested that team processes (e.g., communication) contributed to the development of shared mental models in crews. They found that superior performance was associated with more task-related communications among crew members, specifically more commands, suggestions, statements of intent, exchanges of information, and acknowledgements. They also found that crews with mental models based on shared experiences were able to overcome the effects of fatigue. Based upon these findings, it is expected that:

Hypothesis 3: Operators who talk more to the tether manager will develop better situation awareness through the creation of shared mental models.

Hypothesis 4: More task-related communication between operators and tether managers, i.e., planning of search strategy, reporting of information regarding the environment, and synthesizing of robot-transmitted information with prior knowledge, will enhance the development of shared mental models, leading to better operator situation awareness.

These hypotheses are particularly interesting for autonomous robots because any experiments to confirm them will yield information as to the content of the shared mental models. This would have important implications for the design of training in robot-assisted tasks.

3) SA and Task Performance: The fifth hypothesis completes the link between SA and task performance, building on the previous four hypotheses. To begin the linkage, shared mental models have been positively linked with team performance. In a study of dyadic teams performing a complex computer task, [31] found that shared mental models were predictive of performance. Similarly, in a study of dyadic teams performing a computer-based flight-combat simulation, [32] distinguished between task- and team-based mental models, and found that both related positively to team process and performance. Finally, [33] found that teams that engaged in high-quality planning were able to form a greater shared mental model of each team member’s informational requirements, to pass information to each other in advance of explicit requests for this information during periods of high workload, and to make fewer errors during high workload periods. Effective planning and communication strategies were found to increase team shared mental models and correspondingly team performance. Given these links, it is expected that development of shared mental models will enhance operator SA: therefore it follows that operators will reap the benefits of this heightened SA in terms of task performance. Specifically, it is hypothesized that:

Hypothesis 5a: Operators with better situation awareness will exhibit more effective task performance (structural evaluation, navigation, victim search); and

Hypothesis 5b: Operators with better situation awareness will be more likely to successfully locate a victim during the search process.

These hypotheses directly link SA with task performance in robot-assisted search; confirmation of these hypotheses would have serious impact on the field of rescue robotics, and by extension, other human-robot team tasks as well.

IV. METHOD

This field study was observational and quasi-experimental, i.e., there was a specified setting, task, apparatus and procedure followed. Quasi-experimental experiments differ from "true" or randomized experiments in that experiments assign respondents to treatment conditions at random, while in quasi-experiments, assignment depends on self-selection or administrative decisions (as in this study) to determine who is to be exposed to a treatment condition (in this case, working with a robot). This section describes the data collection procedure, location, and equipment, then the data analysis protocol using the RASAR-CS, and finally the additional measures used in the study outside of the communication analysis.

As described below, twenty-eight operators were observed and videotaped using an Inuktun Variable Geometry Tracked Vehicle (VGTV) robot in a planned, controlled task scenario during a 24-hour disaster response training drill. This provided a repeatable scenario not available in prior studies. Hypotheses 1-4 were addressed through communication analysis using the Robot-Assisted Search and Rescue Coding Scheme (RASAR-CS). Each statement was coded into categories detailing *who* the operators were talking to, *how* their statements were phrased, *what* they were talking about, and *why*. Hypotheses 5a and 5b were examined by combining the results of this communication analysis with three other measures: two subjective ratings (of SA and task performance) and one objective performance outcome (victim found). Operators were rated on task performance and situation awareness using a 5-point Likert scale by two raters. An additional task performance measure consisted of recording whether or not the operator successfully located the victim mannequin.

A. Data Collection

The study setting was a 3-day disaster response training exercise in Bridgeport, CT, for first responders seeking USAR certification to be eligible to serve on a regional Task Force team. The exercise consisted of 2 days of intensive hands-on training followed by a 24-hour deployment evolution on an actual collapse site. As part of the Technical Search Operations training during the first two days, all students received two hours of awareness-level instruction in rescue robotics conducted by researchers from the Center for Robot-Assisted Search and Rescue. The awareness training course was designed to provide the students with a mental model of how the robot worked, and to provide an opportunity for hands-on experience teleoperating a robot in confined space. Course participants were taught the basic procedures for robot-assisted search using the acronym LOVR: Localize, Observe general surroundings, look specifically for Victims, Report.

For the 24-hour high fidelity deployment evolution, an 8-story public housing apartment building was partially collapsed, creating a live (authentic) disaster site (see Figure 2). The site was not simplified and significant safety hazards were present. Large chunks of concrete walls, tangled rebar, and loose electrical wiring posed the main hazards to people on the piles. Weather and visibility conditions were normal for the locale, with overnight temperatures in the 40's (F) and clear to partly cloudy skies. The robot search task scenario was set up on the third floor in the southeast wing of the building. Figure 3 depicts the layout of the search space. The operator and researcher were stationed in the stairwell (Figure 4). The victim mannequin, located in the kitchen (see Figure 5), was visible through a partially collapsed wall from the living area, but the robot could not gain access through the void space. Instead, the operator had to teleoperate the robot further down the hall through debris to get to the kitchen entrance.

The twenty-eight participants in the study were students in the disaster response training exercise. These participants, arbitrarily chosen by the rescue squad leaders over the course of the 24-hour evolution, were a subset of the 175 students involved in the drill, who can be characterized as first responders (firefighters and emergency medical technicians). The majority of the participants were male (93%), between the ages of 25 and 44 (75%), and had 12 or more years of firefighting experience (53%).

The apparatus used in the study was an Inuktun Micro Variable Geometry Tracked Vehicle (VGTV) robot called Bujold. Bujold's robot system consists of a small, tracked platform equipped with a color CCD camera on a tilt unit and two-way audio through a set of microphones and speakers on the robot and Operator Control Unit. Powered and controlled through a 100-foot tether cord that connects the Operator Control Unit and the robot, the Inuktun Micro VGTV is a polymorphic robot which can change from a flat

Figure 2. A partially collapsed 8-story public housing apartment building served as the site for the 24-hour disaster response training evolution in Bridgeport, CT.

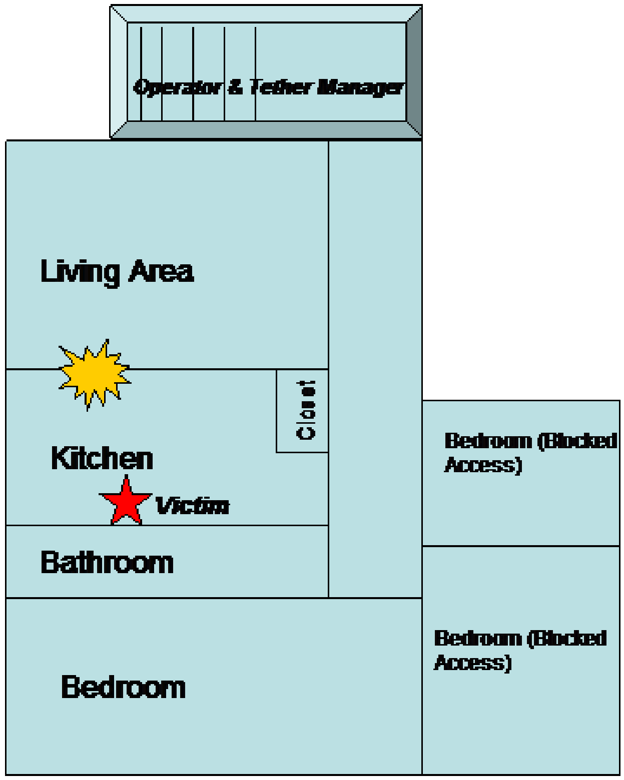

Figure 3. The Robot Operator, Tether Manager, and researcher were located in the stairwell outside the apartment used in the search task scenario.

Figure 4. A robot operator teleoperates Bujold using the Operator Control Unit.

Figure 5. Bujold locates the victim mannequin in the kitchen.

position to a raised, triangular position. Its design allows the vehicle to change shape while moving to meet terrain and obstacle challenges, and it is capable of lifting the camera up to a higher vantage point (about 10.5 inches high when raised to maximum height). The operator is given basic control capability: traversal, power, camera tilt, focus, illumination, and height change for the polymorphic robot.

Human-robot teams (2 people: 1 robot) were videotaped as they performed a technical search task to capture how the robot operator and tether manager used the robot to search for signs of survivors and noted structural information, since these were the activities with direct human-robot interaction. The robot-assisted technical search task is described in detail in [34], and consists primarily of conducting a right-wall following search and structural inspection using the LOVR protocol. Participants were requested to come to the start point in pairs. However, in ten cases a third team member was present (people come and go through the stairwells, some stopped to watch, and some teams had an odd number). The first participant (chosen by consensus between the two participants) acted as robot operator, with the other individual acting as tether manager. The search mission was described by the researcher, and a quick review of the Operator Control Unit was conducted if needed. The first participant searched the designated area until the victim was found or the participant declared the search complete. If the first participant successfully located the victim, the second participant in the pair was instructed to either continue the task, or return the robot to the starting point (depending on the extent of the first participant's search). If the first participant did not discover the victim, the second participant was instructed to continue the search.

The method of data collection was a modified version of the procedure used in [6, 35]. One camera was attached to Bujold's Operator Control Unit to record the view through the robot's camera (what it sees), and a second camera was mounted on a small tripod in the stairwell to simultaneously capture a view of the operator and the Operator Control Unit (what the operator is seeing and doing). A third video unit handheld by one of the researchers recorded an external view of the robot in use when it was possible to do so without being seen by the participants. Sixteen of the 28 runs were in search of a possible victim; the remaining runs were for structural evaluation and navigation back to the start point. In each run, the participants self-organized to use the robot, i.e. they decided who would fill each role, and developed their own strategy given the instruction to search an unsafe area with the robot where a victim was thought to be. The 28 runs yielded a total of 2 hours, 56 minutes of videotape for analysis.

B. Data Analysis, RASAR-CS and Additional Measures

Psychological measurement and analysis techniques were used for this study. This section explains the three steps of the data analysis process: data editing and preparation; data coding and rating; and application of statistical procedures. Next, the coding scheme, or way of reliably categorizing what was said, is detailed. Finally, the three other measures used in this study - a task performance outcome, a task performance rating, and a situation awareness rating - are described.

1) Data Analysis: Data analysis provides a way of organizing the data for the study. The data analysis process begins with data editing and preparation, which time synchronizes the robot's camera videotape with the matching operator videotape to produce a side-by-side video recording of the robot and the operator manipulating the robot. These recordings are then used to code robot movements and statements made by both the operators and surrounding personnel with the Observer Video-Pro [36] behavioral analysis software.

Data coding and rating is the most time-consuming part of the process (in this case, taking 3 man-months: raters must be trained in using the coding and rating systems, a fixed number of statements or behaviors to be coded must be decided upon, and the actual coding requires many man-hours. Raters were two sets of two psychology graduate students who were trained by the first author and another PhD industrial-organizational psychologist to code the videotapes and to provide ratings of SA and task performance for each operator. During 10 hours of rater training, raters reviewed descriptions of the disaster drill and data collection procedures, and then reviewed definitions for all the codes. Behavioral examples selected from other data sets (video recordings) were reviewed, and coding guidelines were developed to reduce ambiguity and to enhance reliability. The majority of the training centered on coding statements together and reaching consensus.

A written transcript of each videotape was produced by research assistants, yielding a fixed number of statements to be coded (1,114 statements across the 28 operators) to be coded. Raters coded each statement across four categories: *dyad* (speaker-recipient pair), *form* (grammatical structure of the communication), *content* (topic) and *function* (intent of the communication). Reliability analyses for the two sets of raters produced Cohen's Kappas of .64 for dyad, .44 for statement form, .27 for statement content and .40 for statement function (n=1,114 statements). Because the reliability indices were lower than desired, raters were asked to conduct consensus ratings, in which rater pairs reviewed the video recordings together, discussed their individual ratings, and came to consensus on each statement. The codes produced by consensus ratings for each of the 1,114 statements were used in data analyses.

Statistical procedures for analysis included quantitative descriptives for the coding categories and measures (means, standard deviations, frequencies and percentages); other quantitative calculations produced were correlational and chi-square analyses, a t-test comparing mean differences in task performance ratings, and an odds-ratio analysis of the *victim found* measure of task performance. Statistical analyses are explained in greater detail in Section V Results. The inclusion of all analyses for all categories is beyond the scope of this article: instead results are reported pertinent to the claims stated in Section I. (Complete analyses are available upon request.) All relationships reported are significant at $p < .05$ unless otherwise noted.

2) RASAR-CS: With the coded statements, the next step was to examine robot operator team interactions using the Robot-Assisted Search and Rescue Coding Scheme (RASAR-CS). RASAR-CS draws on the FAA's Controller-to-Controller Communication and Coordination Taxonomy (C^4T) [37], which uses verbal information to assess team member interaction from communication exchanges in an air traffic control environment. Like the C^4T, the RASAR-CS is domain-specific, capturing not only the "how" and "what" of USAR robot operator teams, but also the "who," as well as observable indicators of robot operator situational awareness.

RASAR-CS addresses the goals of capturing team process and situation awareness in Hypotheses 1-4 by coding each statement on four categories: 1) conversational dyad: speaker-recipient, 2) form: grammatical structure of the communication, 3) content: topic of the communication, and 4) function: intent of the communication. *Team processes* are examined using the dyad, form and content categories to determine which team members are interacting and what they are communicating about. *Operator situation awareness* is explored using elements of the content and function categories. For example, when statements are coded for content, certain elements serve as indicators of the first two levels of situation awareness, perception and comprehension; other elements as indicators of the highest level of situation awareness, planning, projection and problem-solving (see Table 5). Elements in the content and function categories were generated using a Q-sort technique [38], as reported in [39].

Speaker-recipient dyad codes were developed based upon the anticipated roles/individuals present in a USAR environment (Table 1). The primary dyads involve the operator and tether manager (the person manipulating the robot's tether during teleoperation), operator and researcher, or operator and another team member. Nine dyads were constructed to describe conversations between individuals. Five dyad codes classify statements made by the operator to another person (or persons): *operator-tether manager*, *operator- team member*, *operator-researcher*, *operator-group*, or *operator-other*. The *operator-group* dyad is used when the operator is addressing those present as a group, or when the operator's statements are not clearly addressed to a specific individual. The remaining four classify statements received by the operator from another person: *tether manager-operator, team member-operator, researcher-operator*, or *other-operator*. In this study there were only 12 statements (approximately 1%) that included the category element 'other'; therefore these statements were not included in the analysis. Verbalizations between individuals which did not include the operator were not coded.

Elements	Definitions
1. Operator-Tether Manager 2. Tether Manager-Operator	Operator: individual teleoperating the robot Tether manager: individual manipulating the tether and assisting operator with robot
3. Operator-Team Member 4. Team Member-Operator	Team member: one other than the tether manager who is assisting the operator (usually by interpreting)
5. Operator-Researcher 6. Researcher-Operator	Researcher: individual acting as scientist or robot specialist
7. Operator-Other 8. Other-Operator	Other: individual interacting with the operator who is not a tether manager, team member or researcher
9. Operator-Group	Group: set of individuals interacting with the operator

Table 1. Speaker-recipient dyad category elements and definitions used to code who was speaking to whom.

The *form* category (Table 2) describes the grammatical structure of the communication, and contains the elements: *question, instruction, comment* or *answer*. (A statement can be a whole sentence, or a meaningful phrase or sentence fragment.) Statements not matching these categories are classified as undetermined.

Elements	Definitions
1. Question	Request for information
2. Instruction	Direction for task or activity
3. Answer	Response to a question or an instruction
4. Comment	General statement, initiated or responsive, that is not a question, instruction or answer

Table 2. Form category elements and definitions used to code the grammatical structure of communication.

The *content* category describes the topic of communication, and consists of eight elements: 1) statements related to robot functions, parts, errors, or capabilities (*robot state*), 2) statements surrounding the robot's location, spatial orientation in the environment, or position (*robot situatedness*), 3) statements describing characteristics, conditions or events in the search environment (*environment*), 4) statements reflecting associations between current observations and prior observations or knowledge (*information synthesis*), 5) statements concerning the victim (*victim*), 6) indicators of direction of movement or route (*navigation*), 7) statements reflecting search task plans, procedures or decisions (*search strategy*), and finally 8) statements unrelated to the task (*off task*).

Elements	Definitions
1. Robot state	Robot functions, parts, errors, capabilities,etc.
2. Robot situatedness	Robot's location and spatial orientation in the environment; position
3. Environment	Characteristics, conditions or events in the search environment
4. Information synthesis	Connections between current observation and prior observations or knowledge
5. Victim	Pertaining to a victim or possible victim
6. Navigation	Direction of movement or route
7. Search strategy	Search task plans, procedures or decisions
8. Off task	Unrelated or extraneous subject

Table 3. Content category elements and definitions used in coding the topic of communication.

The *function* category is used to classify the purpose of the communication, and is comprised of eight elements: 1) sharing observations about the robot or environment (*report*), 2) projecting future goals or steps to goals (*plan*), 3) asking for information from someone (*seek information*), 4) making a previous statement or observation more precise (*clarify*), 5) affirming a previous statement or observation (*confirm*), 6) expressing doubt, disorientation, or loss of confidence in a state or observation (*voice uncertainty*), 7) sharing information other than that described in *report*, either in response to a question, or offering unsolicited information (*provide information*). In this study, the focus is on the operator's situation awareness; hence an eighth element was included as a default for statements made by individuals other than the operator (*non-operator*). The function elements of *report* and *provide information* merit explanation, as they appear very similar. *Report* involves perception and comprehension of the robot state, robot situatedness, the environment, information synthesis, or the victim. Any other information shared by an operator, in answer to a question or on his own, is classified as *provide information* (e.g., navigation).

Element	Definitions
1. Report	Sharing observations about the robot, environment, or victim
2. Plan	Projecting future goals or steps to goals
3. Seek information	Asking for information from someone
4. Clarify	Making a previous statement or observation more precise
5. Confirm	Affirming a previous statement or observation
6. Voice uncertainty	Expressing doubt, disorientation, or loss of confidence in a state or observation
7. Provide information	Sharing information other than that described in report, either in response to a question, or offering unsolicited information
8. Non-operator	Default for statements made by individuals other than the operator

Table 4. Function category elements and definitions used in coding the purpose of communication.

As mentioned earlier, certain elements in the *content* and *function* categories serve as indicators of specific levels of situation awareness in the coding scheme. Situation awareness is generated through information perceived (Level 1) and comprehended (Level 2) about the robot and environment; the first five *content* elements (*robot state, robot situatedness, environment, information synthesis, victim*) serve as indicators of Levels 1 and 2 SA. Since *navigation* and *search strategy* are elements that cannot be executed efficiently without adequate situation awareness, statements reflecting these are indicators of operator Level 3 SA. Indicators of situation awareness are also captured in the *function* category through the elements *report* and *plan*. When the operator shares information (*reports*) based on the robot's eye view, we can infer the first two levels of SA, perception and comprehension, have taken place. The third SA level, planning and projection, is captured in the function category as the element *plan*.

Category	Element	Levels 1 & 2 SA	Level 3 SA
Content	Robot state	X	
	Robot situatedness	X	
	Environment	X	
	Information synthesis	X	
	Victim	X	
	Navigation		X
	Search strategy		X
Function	Report	X	
	Plan		X

Table 5. Situation awareness (SA) indicators in the RASAR-CS

3) Additional Measures: Because the purpose of this study was to examine the hypothesized link between SA and task performance, three other measures (two of task performance, one of SA) were incorporated to complement the communication analysis conducted using the RASAR-CS. An objective measure of task performance (whether the operator successfully located the victim mannequin) was recorded for 16 of the 28 operators. Raters provided a subjective measure of task performance on a 5-point Likert scale (1=low, 5=high) for all 28 operators. The task being measured depended on the run; as noted earlier, if the first operator in each pair of participants successfully located the victim, the second operator's task consisted of structural evaluation/navigation back to the start point. Raters also provided an overall assessment of each operator's situation awareness during the run, rated on a 5-point Likert scale (1=low, 5=high). These ratings were averaged to produce a mean task performance rating and SA rating for each operator. Intraclass correlation coefficients [40] were used to assess interrater reliability on the task performance and situation awareness measures, producing a reliability estimate of .81 for the task performance measure, and .60 for the situation awareness measure.

V. RESULTS

This section presents results of our analyses pertaining specifically to each of the hypotheses presented earlier. These results are then synthesized into three major findings: 1)

Situation awareness is an important element in robot-assisted tasks; 2) Situation awareness is linked to task performance in robot-assisted tasks; and 3) Shared mental models created through team processes, particularly communication, increase situation awareness.

A. Hypotheses

Strong support was found for Hypotheses 1, 2, and 5, as explicated below. Good support was found for Hypothesis 3, in that the results obtained were consistent with our hypothesis, though not in the manner we expected. Moderate support was found for Hypothesis 4: two of the three types of communication posited to be tied to SA were indeed significantly related.

Hypothesis 1: More operator communications will be related to identification and comprehension (Levels 1 & 2 SA) than to projecting, planning & problem-solving (Level 3 SA) This hypothesis is supported by the statement frequencies and percentages shown in Figure 6. The majority (63%) of operator communications were coded into categories identified as indicators of Levels 1 and 2 SA (*robot state*-26%, *robot situatedness*- 8%, *environment*- 17%, *information synthesis*- 4%, and *victim*- 8%). Indicators of Level 3 SA made up 28% of operator communications (*navigation*-12% and *search strategy*-16%*)*. Operator communications were more about building/maintaining SA than using SA.

Hypothesis 2: Operator communications pertaining to Levels 1 (perception) and 2 (comprehension) SA will be related to communications denoting Level 3 SA (planning, projection and problem-solving). Hypothesis 2 is supported by correlations between operator statements in the *content* and *function* categories (see Figure 7). In the *content* category, operator statements about *search strategy* (Level 3 SA indicator) were significantly related to those about the *environment* and *information synthesis* (r =.78, p<.01, and r = .46), important indicators of Levels 1 and 2 SA. As expected, relationships between the content categories *robot state, robot situatedness,* and *victim* with search strategy were non-significant. (Information about the robot is not pertinent to development of search strategy unless a particular sensor is being discussed; the robot's camera and audio were the only sensors available. Operators don't usually speak of a victim until one is found.) When operators spoke about *navigation* (Level 3 SA indicator), their statements were related to the *environment* (r =.53, p<.01), *information synthesis* (r =.45), and the *robot's situatedness* in that environment (r =.46), all indicators of Levels 1 and 2 SA. Correlations with the content categories *robot state* and *victim* were non-significant. This suggests the robot's position and location in the environment were more important to the operators for navigation than the robot's capabilities or functions. (Statements about the victim were not expected to correlate with navigation.) In the *function category*, indicators of SA are the elements *report* (Levels 1 and 2 SA) and *plan* (Level 3 SA). The correlation between *report* and *plan* (r =.58, p<.01) offers additional support for the importance of perception and comprehension in planning the search strategy.

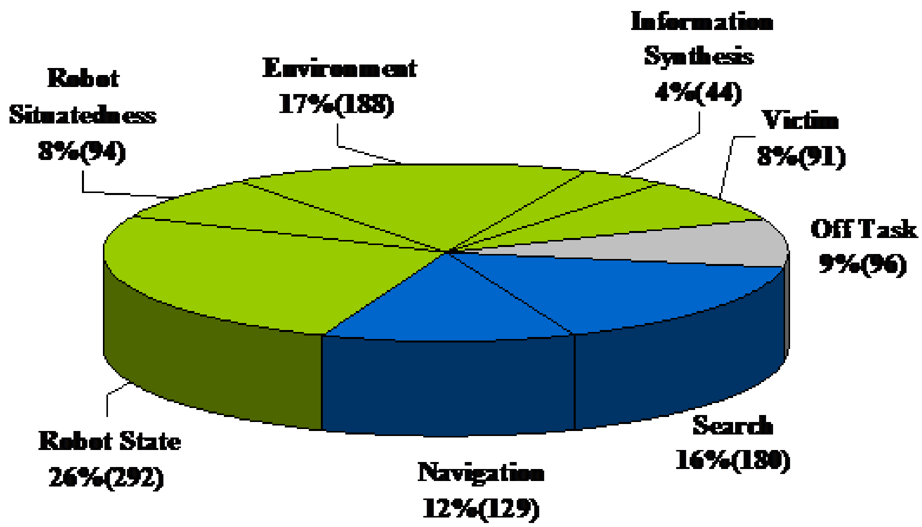

Figure 6. Operator statement content percentages (and statement frequencies in parentheses) by SA level (green = Levels 1 & 2 SA, blue = Level 3 SA).

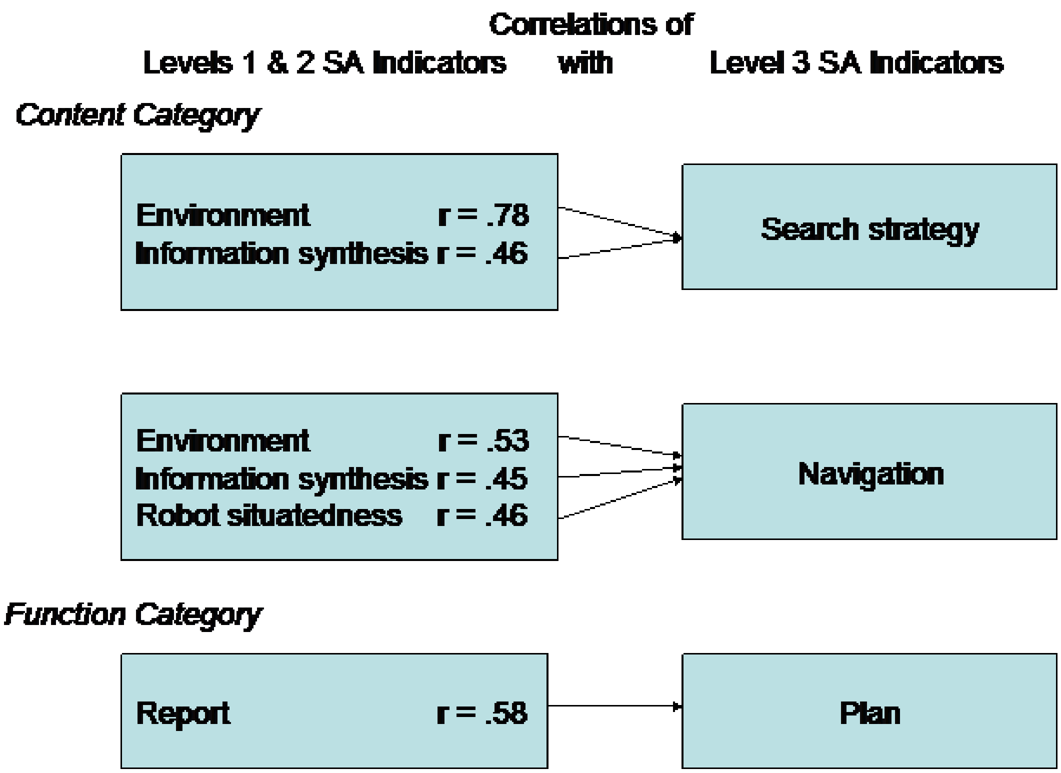

Figure 7. Correlations between Levels 1 and 2 SA (perception and comprehension) and Level 3 SA (projecting, planning and problem-solving) in MOSR teams according to the RASAR-CS.

Hypothesis 3: Operators who talk more to their tether managers will develop better SA through the creation of shared mental models. Statement frequencies and percentages show that for the 28 operators overall, most operator communications (62%) were voiced to the group as a whole rather than to a specific person (see Table 6). Operators directed 13% of their statements to the tether manager, and 7% to a third team member.

Dyad Element	Frequency	Percentage
1. Operator-tether manager	104	13%
2. Operator-team member	49	7%
3. Operator-researcher	132	18%
4. Operator-group	446	62%
Subtotal of statements from operator	731	100% (66% of total)
5. Tether manager-operator	87	23%
6. Team member-operator	105	28%
7. Researcher-operator	191	49%
Subtotal of statements to operator	383	100% (34% of total)
Total statements	1,114	100% of total

Table 6. Statement frequencies and percentages for speaker-recipient dyads.

In chi-square comparisons of high and low SA groups (see Table 7), operators with high SA ratings talked more overall ($\chi^2 = 93$, $p<.0001$) than operators with low SA ratings. They talked more to the group than to specific individuals ($\chi^2 = 124.51$, $p<.0001$). No differences were observed in the frequency of operator communications with the tether manager. However, operators who received low SA ratings had significantly more communications with team members, ($\chi^2 = 12.13$, $p<.0008$), and received more incoming communications from both tether manager ($\chi^2 = 4.9$, $p<.05$) and team member ($\chi^2 = 16.09$, $p<.0001$). These results offer support for our hypothesis; operators did talk to their tether managers more, but addressed them as part of the group rather than individually.

Category	**Element**	**Low SA operators (*n*=11)**	**High SA operators (*n*=17)**	**Chi-square statistic**	***p*-value significance level**
Dyad	1. Operator-tether manager	46	58	1.2	*ns*
	2. Operator-team member	31	18	12.13	.01
	3. Operator-researcher	42	90	2.86	*ns*
	4. Operator-group	59	387	124.51	.01
	5. Tether manager-operator	44	43	4.9	.05
	6. Researcher-operator	75	116	0	*ns*
	7. Team member-operator	61	44	16.09	.01
Content	1. Robot state	119	173	.38	*ns*
	2. Robot situatedness	26	68	5.08	.05
	3. Environment	36	152	31.14	.01
	4. Information synthesis	17	27	0	*ns*
	5. Victim	27	64	3.33	*ns*
	6. Navigation	52	77	.09	*ns*
	7. Search strategy	36	144	26.52	.01
	8. Off task	45	51	2.5	*ns*
Function	1. Report	60	317	84.45	.01
	2. Plan	11	71	22.56	.01

Table 7. Results of chi-square analyses comparing statement frequencies for high and low SA operator groups in selected categories. A large chi-square statistic means there were notable differences between the groups on the item of interest; *p*-values are reported for statistically significant differences. A *p*-value of .05 (or .01) means the likelihood of the results occurring by chance is less than 5% (or 1%) (*ns* = non-significant).

Hypothesis 4: More task-related communication between operators and tether managers, i.e., planning of search strategy, reporting of information regarding the environment, and synthesizing of robot-transmitted information with prior knowledge, will enhance the development of shared mental models, leading to better operator situation awareness. This hypothesis is supported by correlations between statement categories with SA ratings, and by chi-square analyses of high and low SA operators. Significant correlations were found between SA ratings and the operator statement *content* categories of *environment* (r =.41, p<.05) and *search strategy* (r =.49, p<.01); significant correlations were also found between SA ratings and the operator statement *function* categories of *report* (r = .54, p < .01) and *plan* (r = .55, p < .01). Statements related to the *content* category *information synthesis* were not related to SA ratings. High SA operators talked significantly more about the *environment* (χ^2=31.14, p<.01) and *search strategy* (χ^2=26.52, p<.01), and engaged in more *reporting* (χ^2=84.45, p<.01) and *planning* (χ^2=22.56, p<.01). There were no differences in the frequency of operator statements related to *information synthesis* between high and low SA operators, probably due to the low base rate of occurrence. However, high SA operators made more statements related to robot situatedness (χ^2=5.08, p<.05).

Hypothesis 5a: Operators with better situation awareness will exhibit more effective task performance (structural evaluation, navigation, victim search); and

Hypothesis 5b: Operators with better situation awareness will be more likely to successfully locate a victim during the search process.

Hypothesis 5a is supported by the correlation between SA ratings and task performance ratings (r=.81, p<.01), and an independent-samples t-test comparing mean task performance ratings of high and low SA operators. This analysis revealed a significant difference between the two groups, $t(26)$= -4.844, p<.001. The sample means are displayed in Figure 8, which shows that operators in the high SA group scored 26% higher on task performance than did operators in the low SA group (for high SA group, M = 4.06, SD = .659; for low SA group, M = 2.73, SD = .786). Hypothesis 5b is supported by the frequencies and percentages of high and low SA operators who found the victim: 82% of robot operators with high SA ratings successfully located the victim mannequin, while only 60% of the robot operators with low SA ratings did so (Table 8). This translates to an odds-ratio of 9:1 when comparing operators' likelihood to find a victim—operators with good SA are 9 times more likely to find the victim in this search scenario.

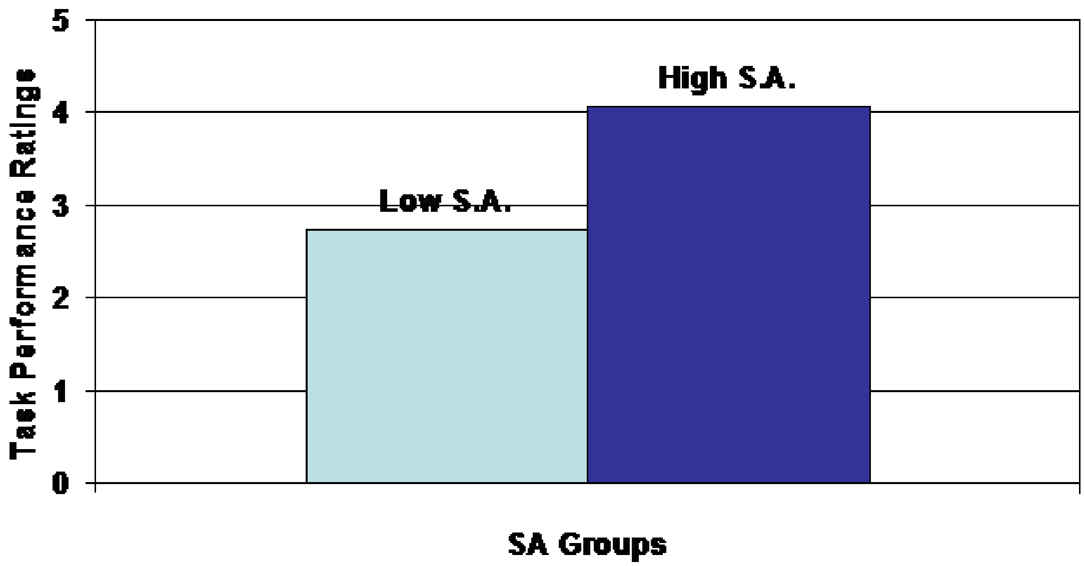

Figure 8. In a *t*-test comparing task performance ratings of high and low SA robot operators, operators in the high SA group scored significantly higher on task performance than did operators in the low SA group, $t(26)= -4.844$, $p<.001$. (High SA group, M = 4.06, SD = .659; low SA group, M = 2.73, SD = .786).

Operators	Victim found	Victim not found	Total
High SA ratings	9	2	11
Low SA ratings	3	2	5
Total number	12	4	16
Total percentage	75%	25%	100%

Table 8. Task performance results (victim found) for robot operators (n=16).

B. Situation awareness is important in robot-assisted tasks

Building and maintaining situation awareness is the primary cognitive task of the robot operator, as evidenced by the support for Hypotheses 1 and 2. The observations of the Bridgeport operators (see Figure 6) revealed that perception and comprehension are indeed more than half the battle in robot-assisted search— nearly two-thirds (63%) of operator statements were related to *gaining* situation awareness; only 28% of the operators' statements related to cognitive tasks *requiring* situation awareness. These percentages are greater than those observed in the Miami study [6] using 5 robot operators. Moreover, operators must build and maintain situation awareness in order to perform the higher level cognitive tasks of devising search strategies and navigation, as evidenced by the correlations between the indicators of Levels 1-2 SA with Level 3 SA indicators (Figure 7). Operators must constantly take in new information about the environment and integrate it with what they have previously observed or known to plan search strategy and navigation; the robot offers very little in the way of assistance on this front, serving primarily as a visual extension of the operator.

C. Situation awareness is linked to task performance

The link between better situation awareness and better task performance in robot-assisted search (Hypotheses 5a & b) is clearly shown in three ways. First, there was a strong correlation between SA ratings and task performance ratings. Task performance ratings for the 28 operators were significantly correlated with situation awareness ratings in the expected direction (r=.81, p<.01). Second, there were sizeable differences in task performance ratings between high and low SA groups. The mean difference in task performance ratings between high and low SA operators was statistically significant as well, highlighting the necessity of situation awareness in all tasks performed by the operators, i.e., structural evaluation, navigation and searching for victims. Third, there was a much higher percentage of successful victim location by high than by low SA operators. The "victim found" outcome measure recorded for 16 operators (Table 8) revealed that, of the high situation awareness operators, 82% were successful in finding the victim, while only 60% of the low situation awareness operators did so. This measure provides the most compelling evidence of the link between situation awareness and task performance in robot-assisted search: the odds-ratio comparing the odds of a high situation awareness operator locating the victim to the odds of a low situation awareness operator locating the victim is 9:1, meaning an operator with good situation awareness is 9 times more likely to successfully locate the victim than an operator with poor situation awareness.

However, the relationship between situation awareness and task performance is not absolute, i.e. there were some high situation awareness operators who did not find the victim, and some low situation awareness operators who did. Two of the three operators who were rated as having low SA, yet found the victim, may have benefited from help and instruction from participants who had just completed the first part of the search and basically pre-empted the role of problem-holder. The third operator may have been distracted: he actually had the victim in sight for several minutes before identifying him, but was making a lot of jokes (off-task) with the group while the victim was in view. The two operators who were rated as having high SA, yet did not find the victim, exhibited signs of the cognitive fatigue that is a factor in any USAR operation [2]. One operator conducted a very thorough search of every room, identifying them (bathroom, closet, bedroom) and describing structural details – yet never mentioned the existence of a kitchen or the need to search for one (he had passed the opening to the kitchen on the opposite side of the hall and never saw it). The other operator evidently could not distinguish the mannequin in the kitchen, though he did report other observations in that area such as the windows and the cabinets. These critical incidents of human error again highlight the need for perceptual assistance and cognitive augmentation.

D. Shared mental models increase situation awareness

Communication between the operator and other teammates involved in the task improves the quality of the human-robot interaction taking place. The benefits of these team processes are evident in the correlations between situation awareness and task performance ratings with various operator-dyad statement categories. *Operator-group* statements were significantly related to both situation awareness and task performance ratings (r = .56, .45 respectively). *Operator-tether manager* communications did not statistically affect operator performance, but the relationships between those statements and the critical *content* areas illustrate the importance of the interplay between the roles of operator and tether manager, and the content of their shared mental model. For example, operator communications with the tether manager were mainly *comments* (r =.54, p<.01) about the *environment* (r =.39), *information synthesis* (r =.61, p<.01), *navigation* (r =.41) and *search strategy* (r =.48). The *function* category elements of *reporting* and *planning* were significantly related to *operator-tether manager* communications (r = .51 and .39, respectively). When the tether manager made comments back to the operator, most were related to *information synthesis* (*r* =.55, *p*<.01), *navigation* (*r* =.68, *p*<.01), and *reporting* (*r* =.49, *p*<.01). In other words, most of the communication between the operator and tether manager revolved around building and maintaining situation awareness, as they pooled their observations to form a shared mental model of the task at hand (see Figure 9). To see this, consider the overall flow of activities and the general model of SA in Figure 1. The robot serves as a conveyor of valuable information not otherwise available to the operator. However, because of the discrepancy between the robot-conveyed information and the operator's usual flow of sensory data (the keyhole effect)[41], he really needs to be able to talk about it and gather more information/interpretation from other perspectives in order to fully utilize the robot's information. Operators with good situation awareness talked out loud to the group as a whole, offering interpretation of the environment, verbally synthesizing what they were seeing with what they already knew about both the task and the situation, and also vocalizing their intentions and plans regarding the search and navigation aspects of the task—in essence, creating the basis of a shared mental model for the group. Like the crew captains in [25], this creation of a shared mental model enabled other team members (in this case, the tether manager) to understand his plans and information needs, and to respond in turn by filling in the informational gaps where possible, and by performing their role in the mission task to support that shared model.

In contrast, communication between the operator and a third team member actually proved deleterious to operator SA and task performance. When a third team member was present in this scenario, he was more of a bystander than an active participant, and may have actually distracted the operator by asking questions about the robot, as evidenced by the negative correlations between *operator-team member* communications and both situation awareness and task performance ratings (*r* = -.52 and -.42, respectively). The contrast is striking: when the operator was talking aloud to the group, the tether manager was able to hear everything the operator said about search strategy, planning, etc. and could thus supply the operator with information on the fly (i.e. before the operator asked for it); and could contribute pertinent information about the robot or the environment from his perspective; the third team member, on the other hand, had no role or vested interest in the outcome, and therefore was "out-of-the-loop" in terms of contributing to the shared mental model held by the others. This is contrary to what we observed in [6]: when a third team member was present, he or she was invested in the search role, looking over the operator's shoulder and assisting in the search process. By contrast, the study conducted in Bridgeport was designed for robot operator-tether manager pairs working together. When a third team member showed up, he was there strictly as an observer, or bystander, and seemed to distract operators overall. The differences in team member roles over the two studies highlight the importance of shared (team) mental models. It remains to be seen, however, whether having a third team member involved in the search task is necessarily a hindrance, or just an artifact of this particular study design.

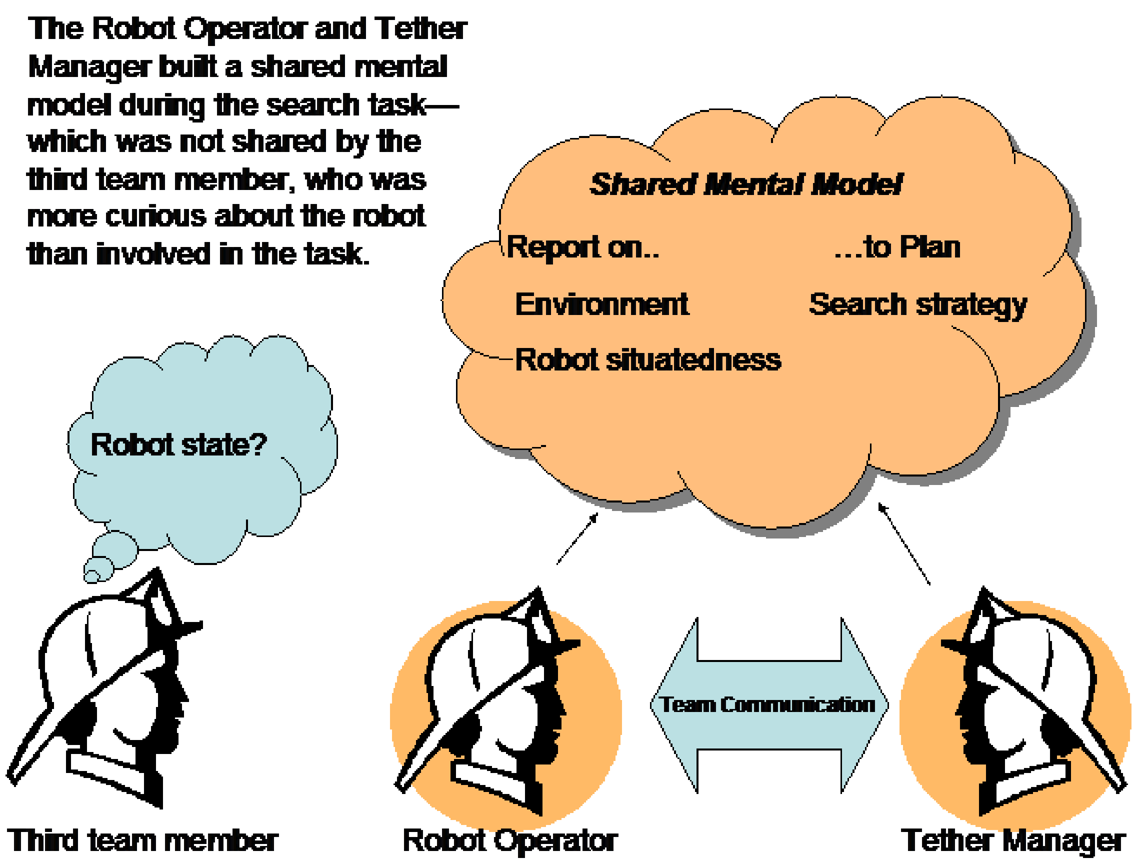

Figure 9. Shared mental models created through team processes and communication increase situation awareness.

VI. DISCUSSION

Although the previous section discussed the three major findings, the study had two broader goals. First is whether the study validated the proposed model of SA formation, with the contents of the robot operator teams' shared mental model being the second. The degree to which these goals have been met is discussed below. The limitations of the study, along with its ramifications for training and cognitive augmentation are then noted.

A. Validation of the Model

The results obtained through communication analysis and other measures offer substantial support for the proposed model presented in Figure 1 describing the formation of situation awareness in robot-assisted technical search, and its subsequent influence on task performance. The validation comes from the development and confirmation of the five hypotheses stated in Section III. In Section V, Hypotheses 1 and 2 showed how the operator develops a role-specific mental model and a situation model based on data from the robot. Hypotheses 3 and 4 explored team communication as the conduit for the development of a shared mental model between the Robot Operator and the Tether Manager, and identified the contents of that shared mental model as task-related: conversations between high-performing teams focused on the environment, the robot's situatedness in that environment, and search strategy. Hypotheses 5a and 5b completed the final link in the chain between SA and task performance, as shown by the large correlation between SA and Task performance, and by the superior performance of high SA operators. Taken together, these results offer compelling evidence for the model. Further research with USAR teams is needed to refine and purify what has been distilled in this effort, looking at not only technical search tasks, but other USAR applications as well, e.g., medical assessment and monitoring of victims during the extrication process. Though the methodology used in this field study is domain-specific and tailored to robot-assisted technical search, the model itself is based on common psychological constructs, and therefore may be generalizable to other robot-assisted tasks and domains, particularly those where robots work with humans in team settings. It may also aid our understanding of the broader constructs of shared mental models and team performance, and how a new technology like robotics may influence team processes and team performance outcomes.

B. What's in the Shared Mental Model?

A second goal of the study was to identify the contents of the shared mental model held by Robot Operators and Tether Managers: what's in the shared mental model? The study indicated that the *environment*, *robot's situatedness*, and *search strategy* are definitely a part of the shared mental model but *information synthesis* and *navigation* might not be.

Communication analysis was used to determine what was being said (*content*) and why (*function*) for all operators, and then SA ratings were used to divide operators into high and low SA groups (under the assumption that high SA operators would have a more fully developed mental model). Chi-square analyses revealed that teams with high SA operators did more *reporting* on the *environment* and the *robot's situatedness* in that environment, and more *planning* of *search strategy*. These *content* and *function* elements (with the exception of *robot situatedness*) were also significantly correlated with SA ratings, further corroborating their place in the shared mental model. Though *robot situatedness* did not correlate significantly with SA ratings, its presence as a significant content area distinguishing high SA operators from low SA operators merits its inclusion in the shared mental model.

It is surprising that the content elements *information synthesis* and *navigation* were missing from this shared mental model (neither approached significance in the chi-square or correlational analyses); the low base rate of occurrence for *information synthesis* statements offers a plausible explanation for its non-appearance, but for *navigation* there is no immediate reason that comes to mind. However, both elements were significantly related to operator-tether manager communications (*information synthesis*, $r = .64$, $p < .01$; *navigation*, $r = .41$), suggesting that they were in fact important in developing the shared mental model held by the team. Their non-appearance in the shared mental model may be due to the limitations of the study, discussed below.

C. Limitations of the Study

This study has two limitations: the lack of experimental control that is endemic to field studies, and the less-than-perfect reliabilities reported for the RASAR-CS and other measures. However, neither of these significantly impact the results reported in this study.

Field research is unpredictable, especially in the USAR environment: the researcher cannot expect (or demand) to have perfect control over the experimental conditions and especially the subjects. The experimental task in this study, while conducted in a real-time disaster response environment, was viewed as a "sidebar experience" by some, since the same area was searched over and over. Though the experiment was designed for a 2:1 human-robot ratio, several teams picked up a third member at the spur of the moment. Some operators approached the task very seriously. Others, caught up in the novelty of the technology, may have performed less effectively than they would have using the robot in an actual search and rescue situation. There may have been some diffusion effects, i.e. operators sharing information about the experiment with others (though we are not aware of any actual occurrences). Too, one cannot discount the effects of time and fatigue on the operators, as evidenced by the differences between those who participated in the experiment earlier in the response and those who participated toward the end. These effects, however, are actually important to our understanding of human-robot interaction in this domain, and therefore should not be controlled (even if we could). The balance between experimental control (more easily attained in the lab) and external validity (field studies' main asset) can be maintained through careful planning (and flexibility) in field research design. In this study the measures taken to exercise experimental control (standardized setting, task, apparatus and procedure followed) offset the variations in number of participants per run, while preserving the invaluable external validity provided by the realistic environment and sample population.

In this study, the reliabilities reported for the four coding categories of the RASAR-CS were lower than desired; therefore consensus ratings were produced to ensure operator communications were correctly coded. Reliable, valid measures are a necessary tool in research. Reliability is the consistency of a measurement instrument, and validity means it measures what it was designed to measure. There is always some degree of error in any measurement instrument, and room for improvement. The RASAR-CS was designed specifically to analyze communications between robot rescue team members during the search task. Subsequent versions have been modified to analyze medical reachback tasks using robots [42] and even social interactions with and through the robot, incorporating gestures and body movements [43]. As we learn more about rescue robotics and new tasks and/or forms of interaction emerge, the coding scheme will likely go through successive iterations to capture important information and the transfer of knowledge not only within the robot team, but also with other entities in the USAR environment, e.g. the rescue squad leader, incident commander, structural specialists, medical personnel, and of course, the victim. These refinements should improve the reliability of the RASAR-CS as well. The interrater reliability estimates reported for the subjective measures of situation awareness and task performance used in this study are acceptable for research purposes; however, multiple-item measures may offer more criterion-related validity if these ratings are to be used in decisions affecting selection, training or evaluation of USAR personnel. Future research should include other measures of team process (coordination, backup behaviors, leadership) in addition to situation awareness.

D. Implications for Training and Cognitive Augmentation

The results of this study have two profound implications for training, not only of USAR teams working with robots, but also for other human-robot team endeavors or cognitive augmentation tools that would aid the operators. The first major implication is that clear roles in the robot-assisted search task seem to promote more effective team process. Team mental models may be enhanced by identification/clarification of the roles and activities involved in robot-assisted search. Since it is at least a two-man task, delineation of operator/tether manager activities may facilitate team performance. If a third team member is available to participate in search, she must be attuned to the goals/nature of the task and respond accordingly; i.e., help the operator determine what is salient through the visual channel and develop search strategy by listening and observing, but limiting her involvement to that role (no driving, no joking around). This has implications for training and performance evaluation of future robot operators in USAR; establishing a list of roles, tasks and activities for robot-assisted search and rescue is clearly necessary to develop training programs and assistive "reminders" as well as standards for performance evaluation.

The second major implication for training (and cognitive augmentation or assistance) is that it needs to focus on the *team* context. Though a common assumption in robotics is that one person can operate or work with one or more robots, the observations made in this study regarding the use of team communication to develop shared mental models of the task suggest that *training* people to work with robots (whether it's one robot or an army of robots) should start at the *team* level rather than the individual. Even if the end goal is to have a one-to-one or one-to-many human-robot ratio, the benefits of team training could speed the process by which users become proficient. The extensive psychological literature on expert-novice differences [44-46] suggests that domain-specific knowledge and the ability to frame and understand a problem are important in expert performance: by identifying the content of high performing operators' mental models, this knowledge and skill acquisition may be addressed through specific training practices. It is interesting to note that in Lumelsky's teleoperation studies [17], he concluded training was useless, though no formal training was implemented as a treatment variable in the experiments reported. Future research examining the effect of training on human-robot task performance is needed, particularly in light of Lumelsky's conclusions…it would be a great disservice if robot designers spent all of their time developing robots for "the untrainable" when an informed training design was all that was required.

VII. CONCLUSIONS

The results of the Bridgeport field study show that situation awareness is indeed linked to task performance in multi-operator single robot (MOSR) teams: MOSR teams with high SA operators are 9 times more likely to find victims. More importantly, these results provide initial validation of the model of SA formation illustrated in Figure 1. The following three findings have increased an understanding of what factors contribute to operator situation awareness and resulting task performance:

First, this study shows that rescuer teams with high SA are 9 times more likely to find victims than rescue teams with low SA, and score 26% higher on ratings of task performance. This clearly demonstrates that high-quality situation awareness does contribute to more effective task performance – though the physical and cognitive fatigue experienced by the operator can cause human error despite having good situation awareness. From a theoretical perspective, this further supports arguments for the inclusion of situation awareness on HRI research agendas, since SA is clearly linked to performance. From a domain perspective, better task performance increases the likelihood of finding victims during the search process. Therefore, research into situation awareness could literally be a life-or-death proposition.

Next, in this study operators spent 63% of their time on basic SA Level 1 and 2 activities (perceiving and comprehending), and only 28% on more cognitive SA Level 3 activities (planning, projecting and problem-solving). This replicated the Miami study findings, i.e. rescuers aren't thinking about which way to go next, but rather where they are, and what they are looking at. This suggests that major advances are needed in sensors and especially sensor interpretation to either augment and support human perception, or to automate it.

Finally, this study identifies at least three distinct types of knowledge in the shared mental model held by the Robot Operator and Tether Manager: *the environment*, *the robot's situatedness*, and *search strategy*. These elements were isolated based upon the two statistical analyses supporting Hypothesis 4: their significant correlations to SA ratings, ranging from .41 to .55, and by their frequency of occurrence in high SA operators, based upon chi-square analyses comparing differences between operators with good or poor situation awareness, with significant chi-square statistics ranging from 5.08 to 84.45. This study also confirmed that the verbal communication, both in terms of content and frequency between the high SA robot team members, was critical in constructing in this shared model. Chi-square analyses showed that operators with good SA talked more during the search process, and that their conversation was focused on goal-salient aspects of the task. This finding suggests ways must be found to encourage and facilitate appropriate communication to support productive team processes. This could include training at the team level or the creation of a software agent to assist the operator. Current work uses the RASAR-CS to examine the use of rescue robots in robot-assisted medical reachback (RAMR), which involves remote medical personnel conducting operator- and robot-mediated victim assessment and triage decision making in an urban search and rescue environment [42]. As the research community begins to create a methodology for studying human-robot interaction, we continue our efforts to develop the RASAR-CS as a tool for future research both in the field and perhaps in more formal laboratory settings as well [47]. Future work examining the formation of SA and resulting task performance in robot-assisted technical search focuses on the impact of the context of communication; in particular, do operators and tether managers need to be nearby in order to create a shared mental model of the task? This is an interesting question: as physical advances in robots eliminate the need for two operators to be in the field with the robot and wireless communications improve, it may be possible have the second person (and possibly others) stationed outside of the disaster area or even in another country. In addition, our exploration of the shared mental model held by the Robot Operator and Tether Manager continues: What is their *common ground* [22] or frame of reference? The relationships described in the model of SA formation would seem to map over to other types of robot-assisted tasks, particularly in the professional service-robot sector. It is hoped that other researchers will take advantage of the rich research methodologies available in the field of psychology to advance our knowledge and understanding of human-robot interaction.

ACKNOWLEDGEMENTS

This work was supported in part by DARPA under the Synergistic Cyber-Forces Seedling program (N66001-1074-411D) and the Cognitive Systems Exploratory Effort program (N66001-03-8921), and SAIC, Inc. The authors would like to thank Jean Scholtz and Ron Brachman for their support, Jenn Casper, Mark Micire, Brian Minten, and Bob Younger for their help in collecting the data, Thomas Fincannon for his assistance in editing and transcribing the videotapes, and Rescue Training Associates for providing the test venue.

REFERENCES

[1] J. L. Burke, R. R. Murphy, E. Rogers, V. J. Lumelsky, and J. Scholtz, "Final report for the DARPA/NSF interdisciplinary study on human-robot interaction," *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, vol. 34, pp. 103-112, 2004.

[2] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, vol. 33, pp. 367-385, 2003.

[3] E. Krotkov and J. Blitch, "The Defense Advanced Research Project Agency (DARPA) Tactical Mobile Robotics program," *International Journal of Robotics Research*, vol. 18, pp. 769-776, 1999.

[4] J. Blitch and R. Murphy, "Alternative Scenarios for Military Deployment of Unmanned Ground Vehicles," presented at 2nd Conference on Robotics for Challenging Environments, 1996.

[5] J. Casper and R. Murphy, "Workflow Study on Human-Robot Interaction in USAR," presented at ICRA 2002, 2002.

[6] 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*, vol. 19, pp. 85-116, 2004.

[7] S. Thrun, "Toward a framework for human-robot interaction," *Human-Computer Interaction*, vol. 19, pp. 9-24, 2004.

[8] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Robotics and Autonomous Systems*, vol. 42, pp. 271-281, 2003.

[9] R. O. Ambrose, H. Aldridge, R. S. Askew, R. R. Burridge, W. Bluethmann, M. Diftler, C. Lovchik, D. Magruder, and F. Rehnmark, "Robonaut: NASA's space humanoid," *Intelligent Systems, IEEE [see also IEEE Expert]*, vol. 15, pp. 57-63, 2000.

[10] Y. Endo, D. C. MacKenzie, and R. C. Arkin, "Usability evaluation of high-level user assistance for robot mission specification," *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, vol. 34, pp. 168-180, 2004.

[11] J. M. Riley, D. B. Kaber, and J. V. Draper, "Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation," *Human Factors and Ergonomics in Manufacturing*, vol. 14, pp. 51-67, 2004.

[12] J. Scholtz, J. Young, J. L. Drury, and H. A. Yanco, "Evaluation of human-robot interaction awareness in search and rescue," presented at Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on, 2004.

[13] 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*, vol. 19, pp. 117-149, 2004.

[14] J. Orasanu and E. Salas, "Team decision making in complex environments," in *Decision making in action: Models and methods*, G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, Eds. Norwood, NJ: Ablex Publishing Corporation, 1993, pp. 327-345.

[15] J. L. Marble, D. J. Bruemmer, and D. A. Few, "Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams," presented at IEEE International Conference on Systems, Man and Cybernetics, 2003.

[16] D. Perzanowski, D. Brock, W. Adams, M. Bugajska, A. C. Schultz, J. G. Trafton, S. Blisard, and M. Skubic, "Finding the FOO: a pilot study for a multimodal interface," presented at IEEE International Conference on Systems, Man and Cybernetics, 2003.

[17] V. Lumelsky, "Human-robot interaction and performance in telerobotic systems," presented at Aerospace Conference, 2003. Proceedings. 2003 IEEE, 2003.

[18] J. W. Crandall, C. W. Nielsen, and G. M. A., "Towards Predicting Robot Team Performance," presented at IEEE International Conference on Systems, Man, and Cybernetics,, 2003.

[19] R. R. Murphy, J. G. Blitch, and J. L. Casper, "RoboCup/AAAI Urban Search and Rescue Events: Reality and Competition," *AI Magazine*, pp. 37-42, 2002.

[20] M. R. Endsley, "Design and evaluation for situation awareness enhancement," presented at Human Factors Society 32nd Annual Meeting, Santa Monica, CA, 1988.

[21] T. Sheridan, *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press, 1992.

[22] H. Jones and P. Hinds, "Extreme work groups: Using SWAT teams as a model for coordinating distributed robots," presented at CSCW 2002 Conference on Computer Supported Cooperative Work, New York, 2002.

[23] V. B. Hinsz, "Mental Models of Groups as Social Systems: Considerations of Specification and Assessment," *Small group research*, vol. 26, pp. 200 (34 pages), 1995.

[24] M. R. Endsley, "Situation models: An avenue to the modeling of mental models," presented at 44th Annual Meeting of the Human Factors and Ergonomics Socety, 2000.

[25] J. Orasanu, "Shared mental models and crew decision making," Princeton University, Cognitive Sciences Laboratory, Princeton, NJ Technical Report No. 46, 1990.

[26] D. Sonnenwald and L. Pierce, "Information behavior in dynamic group work contexts: Interwoven situational awareness, dense social networks and contested collaboration in

command and control," *Information Processing and Management*, vol. 36, pp. 461-479, 2000.

[27] H. C. Foushee, J. K. Lauber, M. M. Baetge, and D. B. Acomb, "Crew factors in flight operations: III. The operational significance of exposure to short-haul air transport operations," NASA, Mountain View, CA NASA Technical Memo No. 88322, 1986.

[28] B. G. Kanki, S. Lozito, and H. C. Foushee, "Communication indices of crew coordination.," *Aviation, Space, & Environmental Medicine*, vol. 60, pp. 56-60, 1989.

[29] M. R. Endsley, B. Bolte, and D. G. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. London: Taylor & Francis, 2003.

[30] D. G. Jones and M. R. Endsley, "Sources of situation awareness errors in aviation.," *Aviation, Space, & Environmental Medicine*, vol. 67, pp. 507-512, 1996.

[31] D. J. Gettman, "An investigation of the shared mental model construct in the context of a complex team task.," in *Dissertation Abstracts International: Section B: The Sciences & Engineering*, vol. 62: Univ Microfilms International, 2001, pp. 2101.

[32] J. E. Mathieu, T. S. Heffner, G. F. Goodwin, E. Salas, and J. A. Cannon-Bowers, "The influence of shared mental models on team process and performance," *Journal of Applied Psychology*, vol. 85, pp. 273-283, 2000.

[33] R. J. Stout, J. A. Cannon-Bowers, E. Salas, and D. M. Milanovich, "Planning, shared mental models, and coordinated performance: An empirical link is established," *Human Factors*, vol. 41, pp. 61-71, 1999.

[34] R. R. Murphy, "Human-robot interaction in rescue robotics," *IEEE Systems, Man and Cybernetics Part A, special issue on Human-Robot Interaction*, vol. 34, 2004.

[35] J. Casper, "Human-robot interactions during the robot-assisted Urban Search and Rescue reponse at the World Trade Center," in *Department of Computer Science & Engineering*. Tampa, FL: University of South Florida, 2002.

[36] L. Noldus, R. Trienes, A. Hendriksen, H. Jansen, and R. Jansen, "The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files," *Behavior Research Methods, Instruments, and Computers*, vol. 32, pp. 197-206, 2000.

[37] L. Peterson, L. Bailey, and B. Willems, "Controller-to-controller communication and coordination taxonomy (C^4T)," Department of Transportation, Federal Aviation

Administration, Office of Aerospace Medicine, Washington, D.C. DOT/FAA/AM-01/19, 2001.

[38] J. Sachs, "Using a small sample Q sort to identify item groups," *Psychological Reports*, vol. 86, pp. 1287-1294, 2000.

[39] J. L. Burke, "Moonlight in Miami: A Field Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise," in *Department of Psychology*. Tampa, FL: University of South Florida, 2002.

[40] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, pp. 420-428, 1979.

[41] D. D. Woods and J. C. Watts, "How not to have to navigate through too many displays," in *Handbook of Human-Computer Interaction*, M. G. Helander, T. K. Landauer, and P. Prabhu, Eds., 2nd ed. Amsterdam, the Netherlands: Elsevier, 1997.

[42] D. L. Riddle, R. R. Murphy, and J. L. Burke, "Robot-Assisted Medical Reachback: A New Kind of Technological Mediation," Center for Robot-Assisted Search and Rescue, Tampa, FL, technical report CRASAR-TR2004-21, 2004.

[43] T. Fincannon, L. Barnes, R. R. Murphy, and D. L. Riddle, "Evidence of the Need for Social Intelligence in Rescue Robots," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Sendai, Japan, 2004.

[44] A. D. de Groot, *Thought and choice in chess*, 2nd ed. New York: Mouton, 1965/1978.

[45] W. G. Chase and H. A. Simon, "The mind's eye in chess," in *Visual information processing*, W. G. Chase, Ed. New York: Academic Press, 1973.

[46] E. Salas and G. A. Klein, *Linking Expertise and Naturalistic Decision Making Expertise, Research and Applications*. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc., 2001.

[47] J. L. Burke, R. R. Murphy, D. R. Riddle, and T. Fincannon, "Task Performance Metrics in Human-Robot Interaction: Taking a Systems Approach," presented at Performance Metrics for Intelligent Systems, Gaithersburg, MD, 2004.