Challenges to Grounding in Human-Robot Collaboration: Errors and Miscommunications in Remote Exploration Robotics

Kristen Stubbs Pamela Hinds David Wettergreen

CMU-RI-TR-06-32

July 2006

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

We report a study of a collaborative human-robot system composed of a science team (located in Pittsburgh), an engineering team (located in Chile), and a rover (located in Chile). The project was intended to be analogous to and inform planetary exploration. We performed observations simultaneously at both sites over two weeks as scientists collected data using the rover. We observed problems in perspective-taking and grounding between the science team, the engineering team, and the rover because of geographic distance and different disciplinary perspectives. Due to this, the science team made errors in commanding the rover and in interpreting the data that was returned to them. Our results have implications for the design of collaboration between people and robots.

Contents

Figure 1: Zoë, the “robotic astrobiologist” used in the Life in the Atacama project.

1 Introduction

The use of robots, especially autonomous mobile robots, to support work tasks is expected to rise significantly over the next few decades [23]. There remains, however, a paucity of empirical research on how users form mental models of robots and their capabilities, how users collaborate with these robots, and the factors that contribute to the success or failure of human-robot collaborative activities. As Burke *et al.* point out [4], most human-robot interaction studies are conducted as experiments in controlled settings. There have been relatively few observational studies of people and robots working together in the unstructured “real world.” Our goal in this research is to better understand how errors and misunderstandings occur in human-robot systems.

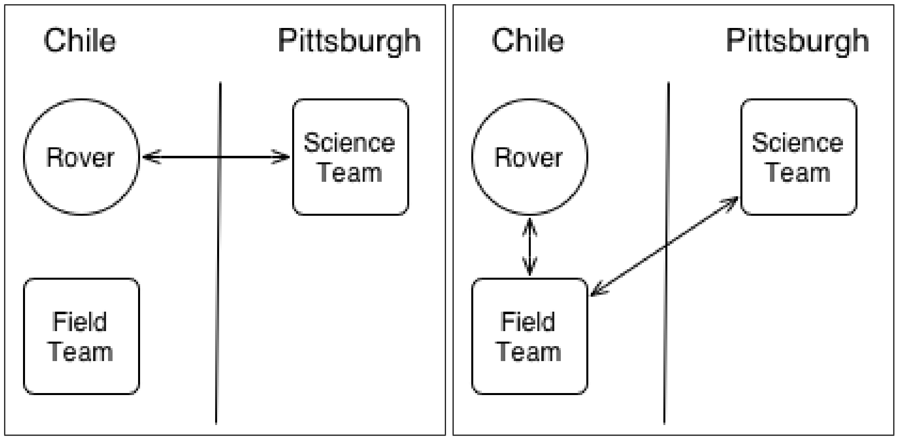
The setting we studied was the “Life in the Atacama” (LITA) project, a project intended to be analogous to planetary exploration but in which the exploration was done on Earth. There were two groups of people and one rover (Figure 1): a group of users commanding the rover remotely and a second group collocated with, monitoring, and often issuing commands directly to the rover. For ease of elocution, we refer to the first group as the science team or broadly as users and to the second group as the engineering team. Our observations of the science team, the engineering team, and the rover showed differences in understandings of the situation. Because of the difficulty of establishing and maintaining common ground between these groups, errors were made in commanding the rover and in interpreting the data that was returned to the science team. In this paper, we explore more deeply the factors that contributed to problems in establishing and maintaining common ground. We also examine the challenges that this presented for the science team's ability to learn and correct during the mission. We discuss implications for the design of human-robot systems.

2 Related Work

Common ground refers to the knowledge, beliefs, goals, and attitudes that people share [9, 15]. Clark and his colleagues [9, 8] propose that common ground is required for successful collaboration—it helps collaborators to know what information is needed by their partners, how to present information so that it is understood, and whether or not the information has been interpreted correctly. At the start of an interaction, collaborators share a certain amount of common ground. For example, if they are members of the same discipline or work group, they likely have a common language and perspective that provides common ground and facilitates communication [14]. Common ground can be developed over time as collaborators share common experiences [7], but it also can be disrupted by factors such as being located in and drawing information from different physical contexts [10]. The interactive process by which common ground is established is referred to as "grounding." Our observations suggest that grounding was problematic for the distributed human-robot teams in our study.

Although the common ground framework was developed to understand conversation and collaboration among people, not between people and machines, recent work has extended the framework into the field of human-computer interaction [2, 17, 21]. This research suggests that interfaces can be improved by thinking about the users' experience as a conversation in which shared meaning between the user and the interface must be developed. By ensuring that common ground can be constructed incrementally, users have more information about what has and has not been understood and can correct accordingly [2].

The common ground framework shares some overlap with work on situation awareness (SA). Endsley[13] defines SA as "knowing what is going on around you." Previous work in SA is largely concerned with whether or not a user has SA, whereas by utilizing the common ground framework, we consider the entire "conversation" that needs to take place between the user and the robot. Situation awareness has recently been examined in the human-robot interaction (HRI) domain, particularly with urban search and rescue (USAR) robots [3, 12, 26]. Casper and Murphy [6], for example, found in their study of human-robot teams responding to the World Trade Center disaster that operators' lack of awareness regarding the state of the robot and how it was situated in the rubble affected the performance of the teams. More recent work indicates that USAR operators spend significantly more time trying to gain SA—assessing the state of the robot and the environment—than they do navigating the robot [3, 12]. Drury et al. [11] have also examined the components of situation awareness relevant for operators of unmanned aerial vehicles. As with work on establishing common ground in human-computer interaction, this work tends to focus on "real time" interaction (with teleoperated robots), so its applicability is not clear for HRI with robots that are remotely and asynchronously commanded.



(a) The science team tries to build common ground with the rover.

(b) The science team's interactions are mediated by the engineering team.

Figure 2: In this study, we examined the grounding process (a) between the science team and the rover and (b) between the science team and the engineering team.

From their observations in the USAR domain, Burke and Murphy [4] propose that shared mental models contribute to SA and that communication is critical to the refinement of shared mental models. They do not, however, test this relationship directly. Our goal in this paper is to examine more closely this process by which users' mental models of robots are formed, updated, and corrected. In other words, we are investigating how grounding occurs or is disrupted between the science team and the engineering team and between the science team and the rover. The science team we observed had difficulty knowing what the rover was doing and what was going on in the environment around the rover. We aim to better understand how the interaction between the science team and the rover (Figure 2(a)) contributed to this confusion and the factors that exacerbated it. The interactions between the science team and the engineering team (Figure 2(b)) are studied to further inform the interactions between the science team and the rover. The common ground framework facilitates such a process focus.

We are aware of several other studies that examine the grounding process as it might be applied to HRI. Jones and Hinds [16] observed SWAT teams and used their findings to inform the design of robot control architectures to coordinate multiple robots. Although their observations did not include robots, their findings established the importance of common ground between a robot and its user, especially when the two are not collocated. More recently, Kiesler and her colleagues [17, 22] have described experiments that report more effective communication between people and robots when common ground is greater. These early studies suggest that the grounding process is critical to effective collaboration between people and robots.

3 Method

The focus of this study was the Life in the Atacama (LITA) project, a multi-site, multi-disciplinary collaboration primarily funded by NASA. The goals of the LITA project are twofold: to use the Atacama desert of Chile as a testing ground to develop technologies and methodologies that may someday be used in the robotic exploration of Mars and to generate new scientific knowledge about the Atacama desert itself. The focus of technology development has centered around a series of semi-autonomous mobile robots and science instrument payloads. Zoë is the most recent rover and was in use during our study (see Figure 1). Zoë is a four-wheeled, solar-powered rover equipped with a number of scientific instruments, including cameras for navigation and for acquiring panoramic images; an on-board near-infrared spectrometer; and an underbelly fluorescence imager used for organofluorescence testing which can detect the presence of biological molecules such as DNA.

For this study, we focused on a particular part of the LITA field season known as remote science operations. Remote science operations involved the use of the rover by two different groups of people: the science team, located in Pittsburgh, and the engineering team, located in the Chilean desert with the rover. The science team was composed of biologists, geologists, and instrument specialists from around the United States and Europe. Their role was to use the rover to search for signs of life in the desert (see Figure 3). The engineering team was composed primarily of roboticists and instrument specialists from Carnegie Mellon University; it also included other instrument specialists and technicians from universities in the United States and Chile. The role of the engineering team was to ensure that the rover was operating safely, to troubleshoot when problems arose, to collect data using instruments that were not yet on-board the rover, and to ensure that the science team was able to gather data successfully (see Figure 4). During our observations, Zoë was a semi-autonomous system under constant development, which required the engineering team to act as an intermediary between the science team and the rover. Thus, the science team sent plans for the rover to the engineering team. The engineering team then interpreted the plans, commanded the rover directly to collect the necessary data, packaged the data, and sent the data back to the science team.

Our observations included a three-day workshop in July 2004, at which time the science team was formally introduced to the rover and the engineering team, and two weeks of remote science operations. During the workshop, the roboticists and instrument specialists presented information about the rover and instrument capabilities and how they would operate. The science team had the opportunity to ask questions, make suggestions, and raise issues about particular features or protocols. The two weeks of remote science operations took place in September and October 2004 with a four week break in-between. During remote science operations, the science team issued daily rover commands from Pittsburgh, PA, to the rover in Chile and received and analyzed data products generated by the rover. One observer conducted observations of the science team in Pittsburgh while one to two other observers simultaneously conducted observations of the engineering team and rover in Chile. The observation process involved writing detailed field notes, drawing diagrams, and taking photographs and video clips. Communication between observers across sites was limited in order to

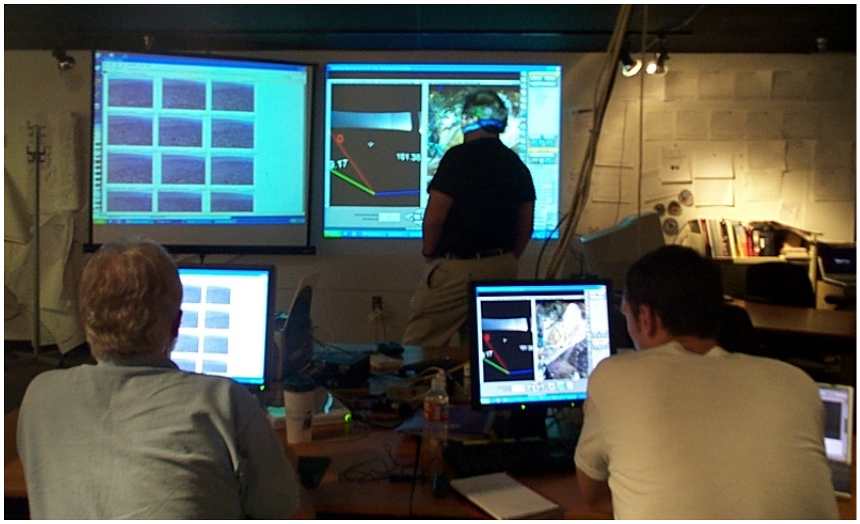


Figure 3: Science team members discuss data returned from the rover.

allow each observer to focus completely on the local situation and to better understand the perspective of the group that she was observing at the time.

All together, 138 hours of observations were conducted in Pittsburgh and 241 hours were conducted in Chile. Field notes, combined with 63 artifact documents, which included PowerPoint presentations from the workshop, emails, and rover plans generated by the science team, formed the data set that was used in our analysis. As the first step in the analysis, several of the authors read the data and identified the high-level issues. The first pass through the data revealed many communication and coordination problems between sites. In the next step, we coded the data based on the problems we observed. Our coding revealed 57 separate common ground problems that occurred during the two weeks of remote science operations. We then used the data to trace the causes of these problems.

4 Results

Although it was evident that both teams were working exceptionally hard to collaborate and to ensure that the mission's objectives were met, our analysis revealed that collaboration was disrupted by two difficulties in the grounding process. The first set of problems were created as a result of geographic separation between the science team and those in the field—the engineering team and the rover. In their work on conversational grounding, Clark and Brennan [7] argue that reliance on communication media impose constraints on the grounding process. In the LITA project, five of the eight constraints mentioned by Clark and Brennan were present, thus making grounding more effortful. Specifically, the science team and those in Chile did not benefit from copresence, visibility (being able to see each other), audibility, cotemporality (receiving

Figure 4: Members of the engineering team work to repair the rover.

utterances at the time they are produced), or simultaneity.

The second source of problems was the science and engineering teams' lack of common ground at the start of the mission. Despite the three-day workshop preceding remote science operations, we found that the science team and the engineering team used different technical languages, had different priorities, and interpreted information differently. Geographic separation and low levels of common ground to begin with made it difficult to establish common ground during the mission. Problems that resulted from these breakdowns then affected the quality of the data that the scientists were able to gather and the quality of (including their confidence in) the interpretations of the data that were returned. We also noticed that the inability of the science team to understand the activities and perspective of the rover made it difficult for them to learn to improve their processes and their ways of interacting with the rover. We used our observations of the interactions between the science and engineering teams to inform our knowledge, more generally, about how people communicate and collaborate with remote robots.

4.1 Problems Resulting from Geographic Separation

We observed a number of problems resulting from geographic separation between the science team and the engineering team and rover. In this paper we will focus on missing contextual information and confusion about the meaning of silence (non-response), two of the most frequent common ground problems that we observed.

4.1.1 Missing contextual information

According to Clark and Brennan, missing contextual information jeopardizes shared understanding because ”the addressee has to imagine appropriate contexts for both the sender and the message” [7, p. 143]. Cramton [10] documented these challenges in her study of graduate students' working on a group project in distributed teams. According to Cramton, difficulties in establishing mutual knowledge arise when team members have difficulty sharing and remembering information about the contexts in which remotely-located collaborators are working. Contextual information includes a wide variety of information that may not be directly related to the task at hand, including anything from national holidays to pressures from supervisors. For our analysis, we consider two types of missing contextual information: information that the science and engineering teams were missing about each other and information that the science team was missing with respect to the rover and its environment.

Missing team information. Through our observations, we discovered that each team was often missing information about the context in which the other team was working. They were therefore not able to take the perspective of their distant colleagues. Perspective-taking can facilitate the grounding process because people understand where their interaction partners "are coming from" can adjust their communication accordingly. The absence of perspective-taking disrupts grounding because the speaker and listener experience and interpret the situation, and therefore what is communicated, differently [18]. In our data, we saw numerous examples of where perspective-taking broke down because information was incomplete or missing. For instance, the science team did not know that the engineering team was constrained by when their cook was able to work. This meant that the engineering team could only eat breakfast at a certain time of the morning, which determined at what time rover operations could begin. The science team, having no knowledge of these constraints, being unaware that they did not know, and feeling strongly about wanting data collected early in the morning, began making negative attributions about the engineering team. After the first phase of science operations, for example, one scientist explained to another that the limitation on morning operations was "because the rover team didn't want to" get up early in the morning.

At the same time, the engineering team was not aware of the difficulties that the science team was having with generating plans for the rover. The user interface for plan generation was difficult and tedious to use. The engineering team, unaware of this, tended to talk about errors in the plan as the fault of the science team personally rather than the result of interface issues or software bugs. For example, after receiving the Day 5 plan from the science team, two engineering team members had a conversation about the actions in the plan being in the wrong order. One engineer said that the science team had the rover traversing to various sites and then returning to the first locale for the day. When asked if that was what the science team really wanted, he replied, "No, they're just confused. Instead of doing panoramas, they are asking us to spin around but not take any pictures."

The engineer in this example was examining a machine-readable plan that had been generated by the plan-generation interface described above. Not being aware of the problems with this interface, nor the fact that the scientists had not directly edited the machine-readable plan, the engineering team's interpretation of the situation included inaccurate characterizations of the science team. Both groups lacked sufficient contextual information to understand what working conditions were like in both locations. This type of missing contextual information could easily be problematic in human-robot teams composed of groups on Earth and astronauts and robots on the moon or Mars. As in the LITA project, scientists or others on Earth may not have access to critical information about the context in which astronauts are working. Such missing information will likely make it difficult to establish common ground and collaborate together effectively.

Missing rover context. So far, our analysis of missing contextual information has focused primarily on perspective-taking and miscommunications between the people who were geographically separated. However, we also saw problems with contextual information that bear on challenges users likely face when interacting directly with a remote robot. Receiving erroneous data from a robot is always a possibility. Without sufficient information about data and the context from which it is collected, making sound scientific judgments can be challenging. The science team received a number of bad data products from Zoë over the course of remote science operations. In one instance, the rover returned two pictures that were supposed to have been taken of its solar panels, but the two pictures were of entirely different parts of the rover. At first, one member of the science team commented that "our targeting's off," but as the team inspected the data more closely, it became apparent that the rover was reporting that the same camera angles had been used for both pictures, meaning that the images should have been essentially identical. In another instance, the science team received a fluorescence image in which nearly half of the field of view appeared to be glowing, signaling the possible presence of life. This caused a great deal of excitement and confusion, as it was unclear whether the team had found a particularly fruitful patch of ground, whether the camera had malfunctioned, or whether there had been some kind of interaction between the dyes used and the ground. After nearly a day spent investigating the mysterious image, the team concluded that sunlight had been shining underneath the rover onto the fluorescence imager resulting in the strange glow they had observed. In both of these cases, a lack of information about the data and its context resulted in confusion and much time spent trying to deduce what could have gone wrong.

Missing contextual information about the state of the rover and the environment in which it was situated was a recurring problem for the LITA science team, especially the lack of information from the rover about the context in which data products were gathered. In most cases, these failures in creating common ground about the contexts that defined the collaboration resulted in errors in data collection or in uncertainty about how to interpret what was collected. This not only wasted valuable time and resources, but it also created frustration at both sites. This is also a potential problem for exploratory robotics missions and USAR tasks, regardless of whether or not any people are collocated with the robots being used.

4.1.2 Interpretations of the meaning of silence

Another common ground problem that Cramton [10] identified in distributed teams was the difficulty of interpreting the meaning of silence. This problem was also prevalent in our data. In distributed teams, the communication media most easily accessible to collaborators (e.g., email, instant messaging, etc.) do not generally support the subtle nuances that people rely on to resolve the meaning of silence. People tend to remain quiet rather than try to resolve problems using these technologies.

In our study, there were several different types of silence that the science team had to struggle to interpret. Most often, silence from the rover came in the form of a missing data product that the science team had requested. A data product could be missing because the rover did not arrive at the specified location, because the necessary instrument temporarily malfunctioned, because the necessary instrument was not functioning at all, because the rover did not have enough energy, because the rover did not have enough time, because the field team had made a mistake in executing a particular protocol, or because the field team chose not to acquire the data at all. In the absence of information about why the data product was missing, the science team was left to speculate about what had gone wrong. Similarly, the science team had to struggle to understand what had happened when the rover had not reached a particular location that they had specified in the plan: Did the rover run out of time or energy? Was the terrain unnavigable? Did the field team simply decide to drive the rover in a different direction? The rover itself provided little easily-accessible information to disambiguate these types of silences, which made it difficult for the science team to understand how to more effectively command the rover and interpret its actions.

4.2 Problems Resulting from Different Disciplinary Perspectives

Besides the common ground problems that arose from geographic separation, we discovered many other problems that stemmed from a lack of initial common ground at the start of the mission. The three-day workshop designed to build common ground between the engineering and science teams proved to be inadequate in bridging the gap between them given initial differences in disciplinary perspectives and the geographic separation of the teams during the mission. A key aspect of the perspective-taking process is estimating what others know [14]. With that information, people can more accurately and efficiently construct messages that will be understood by the person with whom they are communicating. When estimates are inaccurate and common ground is over- (or under-) estimated, communication and collaboration become problematic. In the LITA project, members of the science and engineering teams had different backgrounds and different sets of priorities, and they did not appear to be cognizant of the magnitude of the gap between them. They appeared to fall prey to an egocentric bias—assuming that others' knowledge was similar to their own [14].

Nearly one-third of the problems observed during remote science operations were in part of a result of different disciplinary perspectives between the science and engineering teams. That is, difficulties arose because of the different backgrounds of the individuals in each location. Most of the members of the engineering team were robotics experts with little training in field science; similarly, the science team was composed of experts in biology and geology but not in computer science or robotics. For the most part, the engineering team simply did not understand the full process of doing field science, and the science team did not really understand the process of doing robotics. They also did not understand how important these differences in perspective were given their impact on how the mission would operate.

For example, scientists saw the dew sensor as a very important tool to help them detect subtle environmental changes that could signal the presence of life. In the plan for Day 4, one of the scientists added the following sentence to the plan to emphasize this: "Please install dew sensor on rover ASAP critical." The engineering team did not have the same appreciation for why it was so important for a dew sensor to be physically on the rover, especially given that a dew sensor had already been installed on a weather station a mile away. Because of the two groups' different backgrounds, the engineering team could not fully appreciate why the science team was so insistent on this particular request. Although we observed discussion about the dew sensor among the rover team, it was apparent that this issue was less central to them than it was to the science team. For example, there is no evidence in our data that the rover team told the science team that it was not physically possible for a dew sensor to be mounted on the rover. In conversational grounding, "positive evidence" of understanding is an important element for establishing that both partners in a conversation have a shared understanding of what has been said. Positive evidence might take the form of acknowledgments (e.g. uh huh, yeah, etc.), appropriate responses, or continued attention [7]. In our observations, we saw no positive evidence of understanding regarding the dew sensor, nor did we see attempts to repair this miscommunication. We expect that this was due to the high cost of perspective-taking and repair associated with reliance on media that did not support visibility, audibility, or cotemporality.

The different backgrounds of the engineering team and science team also led to some confusion with respect to the plans that were generated for the rover to execute. Each action in a plan could be assigned a priority, and the science team generally specified these as "high," "medium," or "low." When planning, the science team was focused on getting the rover to a new location so that they could study a new environment. This meant that the actions at the end of the plan were assigned the highest priority. This confused the engineering team because, from their perspective, the actions at the end of the plan were the least likely to happen. Any number of problems could arise that would prevent the rover from getting to the final location specified in the plan: instruments might stop functioning, it might not be physically possible to navigate to the final location, or the rover might run out of energy. The engineering team struggled to make sense of the science team's priorities given the realities of trying to operate a complicated piece of technology in a harsh environment. In a discussion of the Day 2 plan, for example, two engineers talked about the science team's priorities saying that it was odd that their low priority items were first and their high priority items were last. As the conversation continued, one engineer laughed and said that it was backward. The situation continued into the second phase, as evidenced by our observations of R on Day 10:

> At 10:27am, R says, "First, as always, is low priority," and laughs. He

> continues reading the plan and says that the next ones are medium priority and the last one is high priority. R says that he doesn't know how to interpret this. It doesn't make sense to have the first actions lower priority because they don't know if they'll run out of time until later in the day.

The science team remained unaware of many of these challenges of doing field robotics and continued to use a system of priorities that was not clear to the engineering team. As with the dew sensor, we saw few attempts by either team to seek positive evidence of understanding or to take the perspective of the distant team.

This type of problem is described more thoroughly by Weedman [25], who examined the different incentives that technology creators may have from technology users in multi-disciplinary collaborations. Weedman studied a partnership between a group of computer scientists and global change researchers. She discovered that computer scientists' expectations and goals related to the project were very different from the global change researchers' expectations and goals. In order for the global change scientists to advance their own research, they needed quick solutions that could be applied to their current problems. For the computer scientists, however, these quick solutions were not interesting or significant research problems. Thus, Weedman writes that, "At its simplest, the designers' incentive to build the best technical system possible is not well aligned with the incentive for users to move their work directly forward."

In the LITA project, we observed that the issues that were most important to the science team and that shaped the way they developed plans for the rover did not have the same priority for the engineering team; moreover, the teams did not know that these priorities were different. The fact that each team had its own set of priorities and were not aware of the discrepancies affected what information was communicated between the teams and how they interpreted and responded to the information that they received from each other. The best example of this in our data is the case of the bandwidth limitation placed on the science team. According to the design of the mission, the science team was only to be allowed a certain number of megabytes of data per day to be uploaded from the rover. The science team was told this number was 150, but members of the engineering team seemed unconcerned about this limit. The science team went to great lengths to request data products that would allow them to do the most science within the 150 megabytes, going so far as to request lower-resolution or grayscale images in an attempt to save bandwidth. These requests were met with confusion in the field, to the point at which one engineering team member said that they should just tell the science team to convert the higher-resolution images to a lower resolution after the images were transferred. No conversations about the bandwidth limitation were observed in Chile, which supports the argument that it was not a high priority to them. As two engineers discussed (N and O):

> N says that he's not quite sure that to do. What they appear to want, N says, is grayscale images, but they misspecified the file type. O says that they're trying to reduce the size of the files that are being sent. O says he doesn't know why.

The science team never received the grayscale images they wanted from the field, which in turn confused them and made it more difficult for them to generate rover plans.

In human-robot interaction, it is critical that robots' actions reflect users' priorities. As robot autonomy increases, we anticipate that this will be an even more significant challenge as robots use complex schemes to prioritize based on users' requests and on limitations of the robot. Robots will need to verify that they understand the preferences and priorities of users, notice when grounding has not been achieved, adjust their communication to improve the grounding process, and provide users the assurance that a shared understanding has been achieved.

4.3 Effects on Data Quality, Interpretation, and Learning

Our initial analyses suggest that numerous factors interfered with the development of common ground between the science team and the field over the two weeks of science operations. The lack of effective grounding contributed to the science team's not being able to generate realistic plans and to difficulties in interpreting the data returned by the rover. In addition, the science team was not able to learn how to use the rover effectively and take advantage of its capabilities. Theories about conversational grounding suggest that common ground should develop as collaboration occurs. Prior work, however, suggests that establishing common ground may not occur at all when people are geographically separated [10, 18]. In our observations, problems resulting from missing contextual information in turn created problems for data quality, for data interpretation, and for scientists' learning. In one instance, fluorescence images were sent back labeled incorrectly. A particular protocol had been established earlier in the mission that dictated when water and acetic acid were to be sprayed under the instrument and when images were supposed to be taken. The fact that these resulting images were labeled incorrectly caused the scientists to try and figure out what part of the protocol had been executed incorrectly. It was important for the scientists to understand what had happened so that the images could be interpreted correctly. One scientist (L) described the situation as follows:

> L says that "they did it right," but that the labeling indicates that they didn't spray water. L says that if they sprayed it, they didn't record it as something new, just as acetic acid. L says the naming convention says that they should apply the water, then take images, then apply the acetic acid...L says s/he thinks "they skipped that step [taking images after water but before acetic acid]."

In order for L to be able to draw accurate conclusions from the images taken by the fluorescence imager, L needed to know exactly what liquids had been used in what order and at what times images were taken.

Regardless of whether or not people are collocated with the robot, all of the problems we described can impact user's ability to find common ground and learn how to collaborate effectively with a robot. As individuals use a technical system, they develop a "mental model" of how that system operates [5, 24]. In the case of exploration robotics, scientists will then use this model to try to develop an efficient plan for the robot. Some of the factors that may be included in their models include what the robot's capabilities are, how the robot navigates, how the robot estimates priorities and effects trade-offs, and how the robot deploys scientific instruments. Having an accurate understanding of all of these factors is crucial for the science team to construct robot plans that are feasible and return the maximal amount of useful science data.

There were several key problems during LITA remote operations that affected the grounding process and thus scientists' learning. Missing or erroneous data made it difficult for the science team to understand the actual conditions in the field and adapt their exploration strategy to accommodate these conditions. In the example above, a lack of information about what exactly had happened with the fluorescence imager meant that L could not learn about the true behavior of the instrument for use in future planning and data interpretation. In contrast, we also observed some instances in which, with enough information provided to them, the science team was able to learn how to improve the way they used the rover. On the third to last day of operations, for example, the science team was told in an engineering report that the rover had experienced difficulties in crossing a particular section of terrain. The plan for that day had involved the rover's driving in a straight line across what turned out to be very rough terrain. The next day, when the science team wanted to return to a location on the other side of the rough terrain, they were able to change their strategy and create a plan that forced the rover to go around the hazardous area. As a result, the rover was able to navigate back to the previous location successfully. Because of the feedback they received, the science team learned what terrain was easiest for the rover to cross and changed their plans accordingly, which allowed them to collect additional scientific data that they might not have obtained otherwise. In this case, an engineering report (from the engineering team) was the source of information, but this instance makes it clear that information about the rover's context and the interaction between the rover and that context can lead to better perspective-taking, grounding, and perhaps more efficient human-robot collaboration.

The problem of being able to learn how to use a robotic system is applicable to any other domain in which users are naïve, such as scientists in an exploration robotics mission or USAR professionals operating robots in disaster situations. Users are under pressure to make the most efficient use of robots, which will best be facilitated if they can quickly acquire accurate models of how the robots function. Our data show that given sufficient contextual information, users will update their mental models and change their use strategies to successfully meet their goals. However, in the absence of this information, confusion and a lack of confidence in the data can result.

5 Implications for HRI Design

A variety of grounding problems arose during the LITA field season that we observed. Scientists and roboticists were challenged to establish common ground with each other and with the exploration rover, Zoë. We observed problems resulting from geographic separation and from different disciplinary perspectives on the science and engineering teams that resulted in their having different interpretations and priorities during science operations. We argue that many of these problems would be less severe if robots were able to provide more information about what they do and why, especially in the case of failures, and do this in a dynamic way so that incremental grounding can occur. More accurate and timely information about the state of the robot and its actions should help the grounding process and thus enable users to learn how to use the robot more effectively. We recognize that providing such information has been and will continue to be challenging. The need for awareness of the robot and its environment has been explored within the context of SA, especially in terms of the user interfaces provided to USAR operators [26]. Our results suggest that additional contextual information should be provided to users to help in the grounding process:

- Information about the current capabilities of the robot, including technical information about the health of the robot and its instruments. This is typically available in log data but not in a format that is accessible or meaningful to users.
- Status reports about the activities requested by the users. For example, for each data product requested, the robot should provide information about whether that data product was collected, any discrepancies between the request of the users and the location, time, or method used to collect the data, and, if the data product was not collected, why not.
- When failures occur, specific information about exactly what failed and why. This information should be easily accessible to users and provided using terminology and language that are easily understood by users.
- Information relevant to the constraints under which the robot is operating. For example, specific data from the robot about the time, bandwidth, and energy required to collect and transmit each data product.

More specifically, robots need to have a way of figuring out what is known by those who command them and dynamically adjusting what is conveyed. When the science team asks the same questions repeatedly, expresses confusion, or conveys frustration, robots need to detect these cues and adjust their communication accordingly, for example, by providing a more detailed explanation that closes the gap in knowledge. Achieving this is technically challenging due to the need for complex user models and robust representations of users' goals; however, doing so, we believe, will enable the continual adjustment required for collaboration in dynamic environments.

6 Discussion

To date, there is a paucity of research reporting how people collaborate with remote robots and the factors that determine the success of these collaborations. The research we report here establishes a link between grounding problems and errors in commanding a remote robot and in interpreting the data it returns. We contribute to understanding human-robot collaboration, and to CSCW more broadly, in several ways. First, most HRI work has focused on real-time collaboration between people and robots; our work explores interactions that emerge when communication is asynchronous and infrequent. As suggested by Clark and Brennan [7], lack of copresence, visibility, audibility, and, particularly, cotemporality made it difficult for the teams we observed to identify and repair miscommunications. This finding is consistent with research on collaboration among distant people, but we attempt to push beyond previous work into situations in which people work with remote, intelligent instruments [1]. As a result, we hope, more broadly, to contribute insights to the design of better interfaces for collaborating with remote instruments, particularly to the extent that the instruments possess autonomous capabilities.

Second, previous work on situation awareness (SA) is largely concerned with just one aspect of HRI (whether or not the user has SA), whereas by utilizing the common ground framework, we consider the entire "conversation" that needs to take place between the user and the robot. We are therefore able to look more deeply at SA problems by looking at how common ground is created (and disrupted) and propose a more dynamic model than is suggested by SA.

Third, methodologically we demonstrate the value of conducting simultaneous observations at both sites. This parallel observation made it possible to identify more precisely how mental models of the rover, communication between sites, and activities at each site were in or out of sync. We advocate this as a model for CSCW research in situations in which people are working with remote co-workers and/or instruments.

Practically, our results apply directly to several human-robot interaction scenarios. Scenarios like ours, in which some team members are collocated with a robot and others are not, is one that NASA has envisioned for future human and robot explorations of Mars [20]. Robots may arrive first on Mars, directed by humans on Earth to construct basic infrastructure. Once astronauts arrive, the robots may work side-by-side with the astronauts or in combination with people on Earth or at base camp. Urban search and rescue (USAR) teams are also frequently placed in similar situations: robotics experts may operate robots working within a disaster area; however, a person with expertise in USAR but not in robotics is the critical decision maker in the system [3, 19] and may be distant from the scene [3].

Our analysis indicates that finding ways to develop common ground between users and both the robot and the robotic experts (acting as intermediaries between the users and the robot) will be crucial to users ability to collaborate effectively with robots, achieve their goals with minimal frustration, and be confident in the conclusions they draw from data collected by robots on their behalf.

Acknowledgments

The authors would like to thank Nathalie Cabrol, Roxana Wales, and all of the members of the LITA project for their help and participation. This research was supported in part by an NSF Graduate Fellowship to the first author, NSF Grant #ITR/PE-0121426 to the second author, and the NASA ASTEP program under grant NAG5-12890.

References

[1] N. Bos, A. Zimmerman, D. Cooney, J. S. Olson, E. Dahl, and J. Yerkie. From shared databases to communities of practice: A taxonomy of collaboratories. In *Science on the Internet*. In press.

[2] S. E. Brennan and E. A. Hulteen. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, 8:143–151, 1995.

[3] J. L. Burke, R. L. Murphy, M. D. Coovert, and D. L. Riddle. Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human-Computer Interaction*, 19(1–2):85–116, 2004.

[4] J. L. Burke and R. R. Murphy. Situation awareness, team communication, and task performance in robot-assisted technical search: Bujold goes to Bridgeport. Technical Report CRASAR-TR2004-23, University of South Florida, 2004.

[5] J. M. Carroll and J. R. Olson. Mental models in human-computer interaction. In M. Helander, editor, *Handbook of Human-Computer Interaction*, pages 45–65. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.

[6] 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*, 33:367–385, 2003.

[7] H. Clark and S. Brennan. Grounding in communication. In L. B. Resnick, R. M. Levine, and S. D. Teasley, editors, *Perspectives on socially shared cognition*, pages 127–149. APA, 1991.

[8] H. Clark and C. Marshall. Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, and I. A. Sag, editors, *Elements of discourse understanding*, pages 10–63. Cambridge University Press, 1981.

[9] H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[10] C. D. Cramton. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science*, 12(3):346–371, 2001.

[11] J. L. Drury, L. Riek, and N. Rackliffe. A decomposition of UAV-related situation awareness. In *Proceedings of the First Annual Conference on Human-Robot Interaction*, pages 89–94, March 2006.

[12] J. L. Drury, J. Scholtz, and H. A. Yanco. Awareness in human-robot interactions. In *International Conference on Systems, Man and Cybernetics 2003*, volume 1, pages 912–918, October 2003.

[13] M. R. Endsley. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland, editors, *Situation Awareness: Analysis and Measurement*, chapter 1, pages 1–32. Lawrence Earlbaum, 2000.

[14] S. R. Fussell and R. M. Krauss. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 61:378–391, 1992.

[15] S. R. Fussell, L. Setlock, J. Yang, J. Ou, E. Mauer, and A. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19:273–309, 2004.

[16] H. Jones and P. Hinds. Extreme work teams: Using SWAT teams as a model for coordinating distributed robots. In *Proceedings of Computer Supported Cooperative Work 2002, New Orleans, Louisiana*, pages 372–380, November 2002.

[17] S. Kiesler. Fostering common ground in human-robot interaction. In *Proceedings of the IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN)*, pages 729–734, 2005.

[18] R. M. Krauss and S. R. Fussell. Social psychological approaches to the study of communication. In E. T. Higgins and A. Kruglanski, editors, *Social psychology: Handbook of basic principles*. Guilford Press, 1996.

[19] R. R. Murphy. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics*, 32(2):138–153, May 2004.

[20] NASA Advanced Planning and Integration Office. *A Roadmap for the Robotic and Human Exploration of Mars*, May 2005. Available at `http://www.hq.nasa.gov/office/apio/pdf/mars/mars_roadmap.pdf` (Visited Mar. 2006).

[21] T. Paek and E. Horvitz. Uncertainty, utility, and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In *Psychological models of communication in collaborative systems: Papers from the AAAI Fall Symposium, November 5-7, North Falmouth, Massachusetts*, pages 85–92, 1999.

[22] A. Powers, A. D. I. Kramer, S. Lim, J. Kuo, S. Lee, and S. Kiesler. Eliciting information from people with a gendered humanoid robot. In *Proceedings of the IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN)*, pages 158–163, 2005.

[23] S. Thrun. Toward a framework for human-robot interaction. *Human-Computer Interaction*, 19:9–24, 2004.

[24] G. C. van der Veer and M. del Carmen Puerta Melguizo. Mental models. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook*, pages 52–80. Lawrence Erlbaum, 2003.

[25] J. Weedman. The structure of incentive: Design and client roles in application-oriented research. *Science, Technology, & Human Values*, 23(3):315–345, 1998.

[26] H. A. Yanco, J. L. Drury, and J. Scholtz. Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human-Computer Interaction*, 19(1–2):117–149, 2004.