Cognitive Evaluation of Human-Robot Systems: A Method for Analyzing Cognitive Change in Human-Robot Systems

Kristen Stubbs
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Email: kstubbs@cmu.edu

Debra Bernstein
and Kevin Crowley
UPCLOSE
University of Pittsburgh
Pittsburgh, PA 15260
Email: {dlb36, crowleyk+}@pitt.edu

Illah Nourbakhsh
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Email: illah@cmu.edu

***Abstract*— To help answer questions about the behavior of participants in human-robot systems, we propose the Cognitive Evaluation of Human-Robot Systems (CEHRS) method based on our work with the Personal Exploration Rover (PER). The CEHRS method consists of six steps: (1) identify all system participants, (2) collect data from all participant groups, including the system's creators, (3) analyze participant data in light of system-wide goals, (4) answer targeted questions about each participant group to determine the flow of knowledge, information, and influence throughout the system, (5) look for inconsistencies in the knowledge and beliefs of different participant groups, and (6) make recommendations for improvement. We offer this comprehensive, human-centered evaluation method as a starting point for future work in understanding cognitive change in human-robot interactions.**

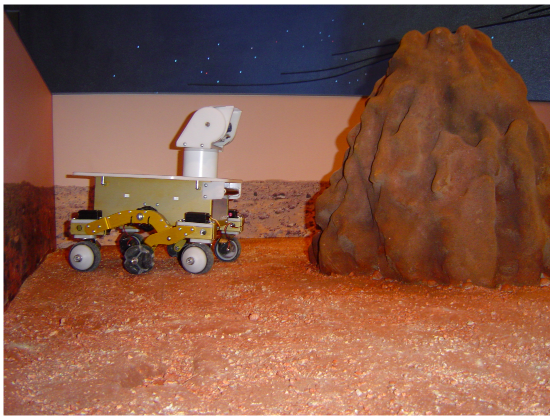

Fig. 1. The Personal Exploration Rover (PER)

I. Introduction

Building better human-robot systems requires that we understand the complex interactions that occur within such systems. As human-robot interaction (HRI) develops, we are becoming more ambitious about the types of interactions we envision for our robots and our users. In particular, we have become interested in the deployment of autonomous robots that are designed to work in complex, real world settings. Our users are not likely to be experts in robotics, and they may possess inaccurate mental models of robotic technologies. Traditional usability studies often treat these human participants as a static factor within the system. However, one of the most interesting aspects of complex human-robot systems is that the human participants in the system will experience cognitive change as they gain more experience with the robot(s). By cognitive change, we mean change in participants' mental models of the system and their understanding of broader concepts within robotics and technology. By mental model, we mean a dynamic representation of a system, a representation which is shaped by, and limited by, the individual's conceptual knowledge [1], [2]. These cognitive changes are emergent and go beyond questions of whether the humans are able to interact efficiently with the robot. Rather, these changes touch upon questions familiar to the learning sciences, questions such as how human mental models are transformed by participation in technology-rich settings. Understanding this cognitive change is crucial to assessing participants' mental models and understanding how and why participants use a system in the way that they do.

To this end, we have developed the Cognitive Evaluation of Human-Robot Systems (CEHRS) method, which facilitates the analysis of cognitive change among different participant groups within a human-robot system. This method is based on our work with the Personal Exploration Rover (PER) (see Figure 1), part of a complex human-robot system [3], [4]. The CEHRS method, which resulted from our studies of various participants in the PER system, aids in the analysis of human-robot systems through:

1) Identification of system participants and examination of their mental models,
2) Comparison of these mental models in light of the relationships between participants in order to determine what information is or is not being transmitted through the system, and
3) Recommendation of improvements to the system which encourage the development of accurate mental models.

This method allows us to identify the areas in which cognitive change has occurred as well as to explain why anticipated or desired changes may have failed to occur.

In this paper, we describe the CEHRS method and demonstrate its application to a specific human-robot system, the PER system. This system was deployed as a museum exhibit at numerous sites around the world, including the Smithsonian National Air and Space Museum, the San Francisco Exploratorium, the NASA/Ames Exploration Center, the National Science Center, and the 2005 World Exhibition in Aichi, Japan. The CEHRS method consists of six steps: (1) identify all system participants, (2) collect data from all participant groups, including the system's creators, (3) analyze participant data in light of system-wide goals, (4) answer targeted questions about each participant group to determine the flow of knowledge, information, and influence throughout the system, (5) look for inconsistencies in the knowledge and beliefs of different participant groups, and (6) make recommendations for improvement. We believe this method will allow system evaluators to identify emergent patterns of use and cognitive change within a system as well as provide the information needed to improve a system's abilities to meet its own goals.

II. Related Approaches

There are many robots that have been designed specifically for use in human-robot interaction studies (see [5] for a survey). The CEHRS method was developed as a result of studying the interactions between various people and one particular type of robot, the PER [3], [4]. Within HRI, a variety of methods have been presented to aid in the analysis of human-robot systems; however, these methods are generally not well-suited to examining the cognitive change that occurs as people use a robot [6].

A. Describing human-robot systems

Several groups have investigated how to classify and describe human-robot systems as a way of analyzing them. Scholtz has proposed a model of HRI based on the roles of supervisor, operator, mechanic, peer, and bystander [7]. When interacting with a robot, people will have initial goals and intentions, execute actions, perceive the results of those actions, and assess the results. Depending on the person's role, the final assessment may result in changes to any of those steps. Scholtz also describes the information that is needed by the person in each role. However, this model does not necessarily represent all participant groups within the system, which is a major focus of the CEHRS method. For example, Scholtz's model hides the role of the creator. Creators' backgrounds and beliefs about the other people who will use a robot have an enormous impact on the system they create. Scholtz's model also does not represent users' previous knowledge, technological background, or attitudes toward technology. All of these things may affect how people use the robot and how they think about it [8].

Yanco and Drury have created a taxonomy which utilizes Scholtz's system of roles and adds ten other attributes which can be used to classify human-robot systems, such as "Task Type," "Robot Morphology," and "Decision Support for Operators" [9], [10]. These attributes capture many interesting features of human-robot systems, but they do not provide a means to represent the cognitive change of operators (or other people in the system).

Arnold *et al.* introduce a more specific type of representation geared towards mixed teams of humans and robots in the domain of space exploration [11]. This representation is primarily focused on the functions and tasks that the system as a whole must perform while also providing a set of "modes" used to describe particular interactions. These modes represent properties of an entire interaction, whether between people or robots, such as "Proximity of Physical Interaction" and "Information Lag." In general, this representation also does not include descriptions of people's knowledge and beliefs; the "Information Exchange" mode is said to characterize "the flow of information between two agents," but this is also restricted to relatively low-level, task-centered information.

All of these ways of describing a human-robot system are useful, but they do not provide the means to represent participants' knowledge and cognitive change.

B. Performance Metrics

Currently proposed metrics for analyzing human-robot systems tend to be primarily task-centered. For example, the metrics being developed by Scholtz *et al.* for use with ordnance disposal robots mainly focus on measures that reflect how efficiently operators can use a robot to perform some number of tasks [12]. Steinfeld *et al.* have proposed a very comprehensive system of metrics for HRI, useful for a wide variety of applications and intended to help "assess the impact of different levels/types of HRI on performance" [13]. In particular, metric 5.2.3, "Accuracy of mental models of device operation," represents what users believe about a particular system, but it does not help us in understanding users' background knowledge or what they learn about more general concepts through use of a robotic system.

C. Methods of Analysis

Kooijmans *et al.* have introduced "interaction debugging" as a new approach to analyzing human-robot interaction [14]. Interaction debugging involves using a comprehensive software tool which integrates data from a robot's sensors as well as external data sources (video cameras, microphones) to monitor how people interact with a robot. A researcher can examine the data frame-by-frame to understand exactly how a user touched, talked with, or otherwise reacted to the robot. This method is undoubtedly useful for examining people's behavior in the presence of a robot, but it may not be sufficient to capture data about people's beliefs and attitudes, nor does it help researchers understand the role of other people in the system, such as creators or maintainers. In addition, this method does not provide any guidance for how to use the data that can be displayed within the tool.

III. Background

A. The Personal Exploration Rover

The CEHRS method was developed as a result of several studies involving the Personal Exploration Rover (PER). The

PER was originally designed as a tool to help educate the public about the NASA Mars Exploration Rover (MER) mission. The PER was envisioned to be reminiscent of the MER, with a camera and an infrared rangefinder mounted in a pan-tilt head, six-wheel rocker-bogie suspension to allow the PER to surmount small obstacles, and an on-board ultraviolet light that is used to look for "signs of life."

The PER is the key component of a museum exhibit which has been installed at various sites around the world. For the exhibit, the PER is placed inside a simulated Martian environment (the "Mars yard"). At a computer kiosk, visitors obtain a real-time panoramic image taken by the PER in conjunction with an overhead map of the Mars yard to specify a direction and distance for the rover to travel. Once this simple, user-directed plan has been sent to the PER, the rover begins moving autonomously according to the specified direction and distance. As it moves, the PER scans its head back and forth, using the infrared rangefinder to detect and avoid obstacles. Once the PER has traveled the specified distance, it begins to scan for the target rock. Upon locating a rock, the PER performs an autonomous close approach, shines its ultraviolet light, and returns a picture of the target to the visitor. On some rocks, fluorescing paint is used to simulate Martian life; visitors receiving a picture of a glowing rock have found "signs of life" on Mars.

B. Motivation

In our previous publications, we examined participant groups' beliefs about robotics and the capabilities of the PER [3], [4]. In the current paper, we examine the system as a whole and use the CEHRS method to identify why some participants did not experience cognitive change with respect to a particular high-level concept salient to the PER system, the concept of rover autonomy. The exhibit's creators intended for the exhibit to promote cognitive change by encouraging visitors to think about the rover's autonomous behaviors and the importance of rover autonomy [3]. However, only about half of the museum employees and fewer than half of the children interviewed described the PER's autonomous capabilities. In the following sections, we will explain the CEHRS method and use it to analyze how and why the concept of rover autonomy was not clearly understood by all participants in the PER system.

IV. CEHRS Method Description

The CEHRS method for analyzing human-robot systems consists of six steps. The first step is to *identify system participants*. Participants include all system users (both intended and incidental) and, importantly, the system's creators.

The second step is to *collect data* from each participant group. As Suchman has argued in [15], it is critical not to make assumptions about the activities of any group. Data collected should include participants' experiences with different aspects of the system as well as their a priori expectations and mental models for how the system works. Collecting this information from the system's creators is equally important. Creators' mental models and expectations shape all aspects of system development thereby influencing the participants' experience. Because learning occurs within a larger context, as Lave has argued in [16], understanding cognitive change requires that we obtain data from all system participants. A number of methods exist for collecting this type of data, including (but not limited to) interviews, surveys, direct observations of participants interacting with the system, and system logs. For systems deployed over an extended period of time, data collection should occur at multiple intervals.

Step Four: Targeted Questions

1) Who are the other members of the system with whom this group interacts; whom else is this group responsible for thinking about?
2) What are the general attitudes, beliefs and knowledge of this group? To what extent does this group share knowledge with experts in the field?
3) Do members of this group understand how general concepts are instantiated in this particular system?

Fig. 2. Three targeted questions provide a focus for analysis.

The goal of step three is to *analyze* the data. At the most basic level, data analysis can confirm that all participant groups are interacting with the system in ways consistent with the system's goals. Data analysis can take many forms depending on the nature of the data collected. Quantitative data can be examined for patterns and trends, which can help to establish how participants used the system. Qualitative data, such as interviews, can be coded to identify recurrent themes in user responses and illuminate user beliefs about the system [17]. This first round of analysis can sometimes yield unexpected results: different participant groups may possess different beliefs about the system.

In step four, we probe these differences as we *answer targeted questions* about each participant group (see Figure 2). The goal of this step is to increase understanding of how the system functions on a cognitive level.

The first question, which focuses on the relationship between participant groups, serves two functions. It identifies central and peripheral (or unintended) user groups and also helps illuminate connections within the system. Connections are important because information and influence can flow through the system at any connection point. Cognitive science research suggests that the degree to which participants in a system think about others directly impacts their actions and contributions [18], [19], making this a relevant consideration in system evaluation. The second and third questions are designed to assess participants' level of knowledge, both global and local to the system. These questions are particularly important to ask in systems that include non-expert users, as expert and non-expert users are likely to have different explanatory models for the same system [20].

The fifth step is to *find mismatches* in participant groups' responses to the three questions stated above. The process begins by examining each group's answer to the first question and identifying any user groups whose needs are not being considered by other groups (or the creators). Next, each group's answers to the second and third questions are compared in light of the system interaction patterns identified in the first question. Inconsistent models among participant groups may suggest that there has not been sufficient cognitive change following initial participation in the system for all participants to attain accurate knowledge.

The last step is to consider the findings and *recommend improvements*. The inconsistent responses identified in the previous step can be used to pinpoint areas in need of improvement. These improvements may include increasing the amount of information that is available to different user groups or re-designing the robot or other aspects of the system in order to accommodate the unmet needs of particular users and/or promote accurate mental models.

V. METHOD APPLICATION

We now apply this method in the context of the PER system in order to understand why the concept of autonomy was not successfully conveyed through the system despite the fact that this was a major goal of the exhibit's creators.

1. Identify system participants. Participants in this system include the PER's creators, the museum employees who work with the PER, and families visiting the PER exhibit. Each of these groups is interconnected, as shown in Figure 3.[1] As shown in the diagram, creators considered how museum employees and visitors would use the system, museum employees maintained the system while assisting visitors in using it, and parents helped their children to use the exhibit and learn from it.

2. Collect data. For each group, we have collected data through interviews and observations. We believe it is imperative that this type of information be collected from all participant groups, as only this level of data allows us to understand the beliefs, misconceptions, and mental models driving each group's participation in the system. In particular, we collected data regarding each group's beliefs about robot autonomy, as autonomy was a key feature of the PER.

The team that created the PER and its exhibit is composed of managers, programmers, interaction designers, and engineers. Six of the individuals who were key in the design, development, and deployment of the PER, representing all of these professions, were interviewed about their work on the project. The resulting data were used to better understand the background of the creators and their goals for the system. Data collected from museum employees and visitors have been published in [4] and [3], respectively.

3. Analyze the data. Any visitor or museum staff member who interacted with the PER had numerous opportunities to see the rover perform autonomous actions, such as avoiding obstacles or performing an autonomous close approach in the Mars yard. Our initial data analysis revealed that only slightly fewer than half of child visitors and slightly over half of the museum staff interviewed recognized the PER's autonomous capabilities [3].

Why did so few users recognize the PER's autonomous capabilities? To be sure, this is an important question for the designers of the PER system, as teaching users about robotic autonomy was one of the goals of the system. However, this question also extends beyond the PER system, as the discussion is relevant for any human-robot system in which non-expert users will be working with an autonomous robot, such as in an office setting (see [22]). Any time a user is engaged with a robot in a goal-directed task, the user's mental model becomes a potentially important tool for guiding his or her behavior. Especially when the user is a novice, there is the potential for the user to form an incorrect mental model of the robot [20]. This is what happened in the case of the PER: some museum visitors and museum employees failed to recognize that the PER was capable of autonomous behavior. Because visitors' interactions with the PER were heavily mediated by an instructive computer interface, these misbeliefs did not affect their ability to use the exhibit or interact with the PER. However, in other circumstances, an incorrect model of the robot could potentially hamper an individual's ability to successfully complete a task [10], [23].

4. Answer targeted questions. These questions about autonomy raised during our initial analysis prompted us to conduct a second, more detailed analysis asking three more focused questions of each of our groups (see Figure 2).

Creators: *1. Whom else are the creators thinking about?* It is clear from the creator's interviews that they thought about the visitors while designing the PER system. When asked to predict what visitors would assume about the PER's capabilities, one creator replied, "If I did not actually work on the PER myself, and I knew nothing about it, and I came and I saw it, I would start looking around for, like, where's the exterior cameras that's keeping track of this thing. I wouldn't think that it was truly autonomous." With this knowledge in mind, the creators designed the PER exhibit to draw visitors' attention to the autonomous features of the robot. For example, when the PER drives around the Mars yard, it turns its head to the right and left to simulate searching behavior. The interface also reinforces the idea of autonomous search by using language such as "scanning terrain to find your target rock". These design features are all central to the PER system goals, which included teaching the public about the importance of on-board rover autonomy.

2. What do the creators know about the concept of autonomy? The creators were clearly fluent with the concept of robotic autonomy and believed that it was an important capability, particularly for space rovers.

3. How do the creators understand autonomy within the PER system specifically? The creators had the most robust understanding of the PER's autonomous capabilities and the greatest ability to explain them. Creators described the robot's ability to "dead-reckon", "scan for obstacles", and "make decisions about when it is clear of an obstacle."

Museum Employees: *1. Whom else are museum em-*

[1] This diagram is based upon the Collaboration Diagram from the Unified Modeling Language [21].

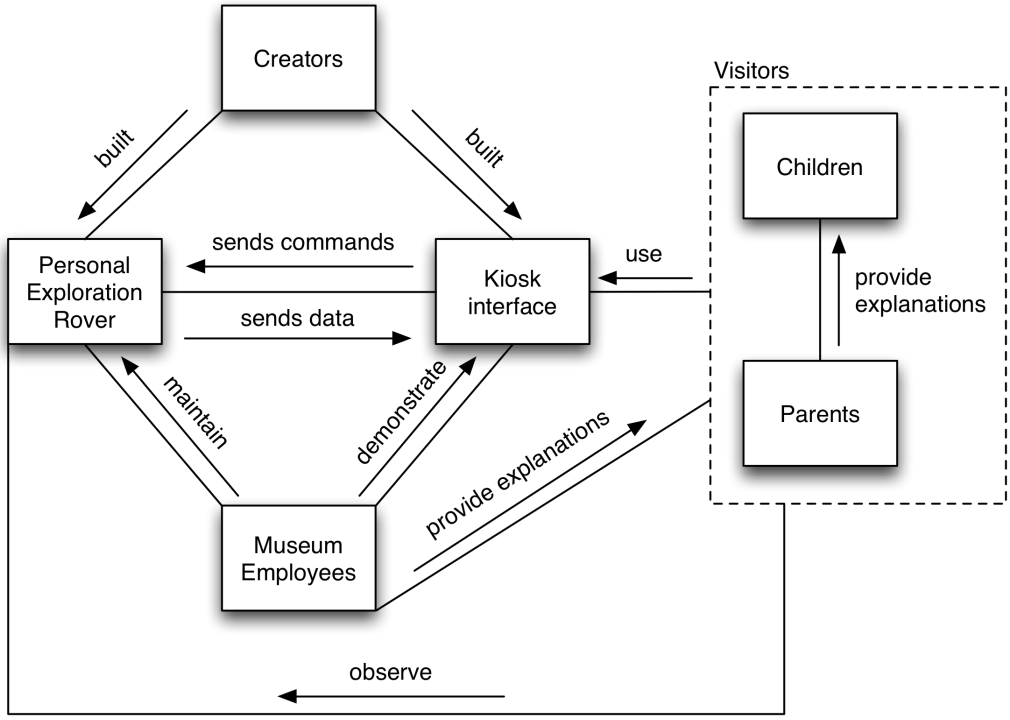

Fig. 3. Collaboration diagram of all members of the PER system and their interactions.

ployees thinking about? The answer to this question was undoubtedly visitors; overall roughly one-third of museum employees' talk was about visitors [4]. Twelve out of 13 museum employees discussed what they felt visitors were learning or what they hoped visitors would learn as a result of using the PER exhibit. Of these, only two mentioned autonomy specifically as something about which visitors would learn; three others described autonomous behaviors such as saying that visitors would learn that the rover is "controlling some of its own motions." All together, fewer than half of the museum employees mentioned robot autonomy as a lesson that visitors would learn by using the PERs.

2. What do museum employees know about the concept of autonomy? Only three employees talked specifically about why it is important for rovers to be autonomous (23%). This suggests that the idea of robot autonomy was not something that was familiar to all employees.

3. How do museum employees understand autonomy within the PER system? Of the six museum employees who explained why they believed the rovers to be intelligent, only one specifically mentioned autonomy although three others talked about how the rover "can make some decisions on its own" and "corrects itself," phrases which describe autonomous behavior. However, two museum employees called the PER intelligent because "you tell it what to do, you tell it where to go." This idea is inconsistent with the creators' ideas of intelligence and autonomy.

Visitors: *1. Whom else are visitors thinking about?* Most visitors were primarily engaged with the robot; however, parents were concerned about their children's understanding of autonomy and tended to talk about it by pointing out specific instances in which the PER demonstrated autonomous behavior. For example, one parent explained the PER's searching behavior by saying, "He just looked around to see if he could find the rock that you wanted him to go to."

2. What do visitors know about the concept of autonomy? The extent to which visitors discussed rover autonomy at the PER exhibit was partially a function of their prior knowledge and interest in the MER mission. This was quantified by calculating an 'interest score' for each family (see [3]). Families with high levels of interest in the MER mission discussed the PER's autonomous capabilities more often than those with low levels of interest [3].

3. How do visitors understand autonomy within the PER system? Visitors exhibited a wide range of knowledge regarding the autonomous capabilities of the PER. Like museum employees, nearly all visitors had the opportunity to observe the PER acting autonomously (e.g., during the close approach). However, prior research suggests that some children do not believe robots are capable of intelligent behavior [24]. It is possible that some visitors will observe the PER and understand that the robot is acting autonomously while other visitors will fail to realize this. More than half of the children observed at the exhibit fell into the latter category. As a 10-year-old child said about the PER during a post-exhibit interview, "...It's kind of like a servant being told what to do. If you tell it to do something, it will do it." Despite having seen the behavior, this child still failed to attribute autonomous action to the PER. Rather, he understood the PER to be solely following the user's directions.

5. Find mismatches. In our system, we did not identify any participant groups whose needs were not considered, so there were no mismatches for the first question. Comparing the answers to the second and third of our three questions across each group reveals a number of mismatches:

- For the second question, creators clearly understood the concept of autonomy; however, museum employees rarely talked about the concept. Only visitors with strong interest in the MER mission tended to discuss autonomy as they used the exhibit.
- For the third question, creators also talked about specific behaviors that the PER uses which demonstrate its autonomy. Some museum employees' descriptions of the PER's behavior were inconsistent with the creators' ideas. In addition, more than half of child visitors also failed to recognize the PER's autonomous capabilities.

6. Recommend improvements. Based on the mismatches discussed in the previous step, it is clear that not all participants experienced cognitive change such that they attained general knowledge about the concept of autonomy and knowledge about how it was instantiated within the PER system. This demonstrates that designing a rover that produces autonomous behavior may not be enough to change users' conceptions. While the PER exhibited several obviously autonomous actions, its design assumed that users were comfortable thinking of robots as autonomous entities. However, prior and current research suggests that naïve users (particularly children) may not think of robots as capable of intelligent behavior. Changing this belief may be a first step towards changing people's beliefs about robotic autonomy.

At the robot and user levels, one way to do this may be to provide more explicit information about what the robot is doing. For example, informing visitors that the PER is no longer relying on user input to complete its task might be useful. Given that the previous study on museum employees demonstrated anthropomorphism to be an important part of employees' models of the PER, is it possible that using anthropomorphic language within the context of the exhibit might help visitors and employees to notice autonomous behavior [4]? Rather than merely reporting its state, messages from the PER to the visitor could emphasize how the PER is acting autonomously ("I see an obstacle in front of me, so I will try to find a way around it"). Another suggestion is to show visitors the difference between how an autonomous and a non-autonomous robot complete a task.

At the system-wide level, we may want to consider the fact that the PER exhibit was designed so that all users could complete a mission, regardless of their misconceptions about the robot. No doubt this is a positive feature in a museum exhibit. However, as a system feature, allowing users to accomplish their goals with incorrect models could potentially hinder the development of correct mental models.

VI. Conclusions

In this paper, we have presented the Cognitive Evaluation of Human-Robot Systems (CEHRS) method for examining cognitive change in human-robot systems. We have applied this method to the PER system to analyze why particular cognitive changes failed to occur among some participant groups. Beyond usability, the CEHRS method allowed us to identify participant groups, examine and compare their mental models, and suggest how the system could better facilitate desired cognitive change. The CEHRS method consists of six major steps:

1) Identify participants.
2) Collect data.
3) Analyze the data.
4) Answer targeted questions.
5) Find mismatches.
6) Recommend improvements.

Because how people think about a robotic system may effect how they interact with it, applying this method may offer benefits to a variety of human-robot systems. Our method encourages examination of and assists in understanding how and why cognitive changes take place. The six steps presented here are necessary for understanding complex interactions between groups of people and robots, but we recognize that they may not be sufficient. We offer the CEHRS method as a multidisciplinary, theoretically- and empirically-grounded starting point for future work aimed at better understanding cognitive change that takes place within complex human-robot systems.

Acknowledgment

This work is partially supported by NASA/Ames Autonomy, Intel Corporation, and an NSF Graduate Student Fellowship to K. Stubbs. We would like to thank Anuja Parikh and Catherine Eberbach for assisting us in data collection as well as all of the staff and visitors of the NASA/Ames Exploration Center, the National Science Center, the San Francisco Exploratorium, and the Smithsonian National Air and Space Museum who participated in this research. We are especially indebted to Marti Louw for her insightful comments and suggestions. We would also like to thank Terry Fong, Emily Hamner, Sara Kiesler, Mark Lotter, Sasha Palmquist, Skip Shelly, and many others who provided valuable feedback.

References

[1] P. N. Johnson-Laird, "Mental models in cognitive science," *Cognitive Science*, vol. 4, pp. 71–115, 1980.
[2] S. Vosniadou and W. F. Brewer, "Mental models of Earth: A study of conceptual change in childhood," *Cognitive psychology*, vol. 24, pp. 535–585, 1992.
[3] I. Nourbakhsh, E. Hamner, D. Bernstein, K. Crowley, E. Porter, T. Hsiu, B. Dunlavey, E. Ayoob, M. Lotter, S. Shelly, A. Parikh, and D. Clancy, "The Personal Exploration Rover: The ground-up design, deployment and educational evaluation of an educational robot for unmediated informal learning sites," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-04-38, December 2004.
[4] K. Stubbs, D. Bernstein, K. Crowley, and I. Nourbakhsh, "Long-term human-robot interaction: The Personal Exploration Rover and museum docents," in *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, July 2005, pp. 621–628.
[5] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots: Concepts, design, and applications," Carnegie Mellon University, Tech. Rep. CMU-RI-TR-02-29, December 2002.
[6] D. Norman, "How might humans interact with robots?" 2001, essay forming the basis of keynote address to a DARPA/NSF workshop on Human-Robot Interaction. Available at `http://www.jnd.org/dn.mss/how_might_humans_int.html` (Visited Sept. 2005).

[7] J. Scholtz, "Theory and evaluation of human robot interactions," in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, January 2003, pp. 125–134.

[8] 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)*, 2005, pp. 158–163.

[9] H. A. Yanco and J. L. Drury, "A taxonomy for human-robot interaction," in *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction, AAAI Technical Report FS-02-03, Falmouth, Massachusetts*, November 2002, pp. 111–119.

[10] 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, no. 1–2, pp. 117–149, 2004.

[11] J. A. Arnold, J. H. Saleh, and J. A. Hoffman, "A formal representation for heterogeneous teams of humans and robots in space exploration," in *AIAA-2005-6748, AIAA Space 2005*, September 2005.

[12] J. Scholtz, M. Theofanos, and B. Antonishek, "Development of a test bed for evaluating human-robot performance for explosive ordnance disposal robots," in *2006 Human-Robot Interaction Conference*. ACM, March 2006.

[13] A. M. Steinfeld, T. W. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the First Annual Conference on Human-Robot Interaction*. ACM, March 2006.

[14] T. Kooijmans, T. Kanda, C. Bartneck, H. Ishiguro, and N. Hagita, "Interaction debugging: an integral approach to analyze human-robot interaction," in *2006 Human-Robot Interaction Conference*. ACM, March 2006.

[15] L. Suchman, "Making work visible," *Communications of the ACM*, vol. 38, pp. 56–64, September 1995.

[16] J. Lave, "The practice of learning," in *Understanding practice: Perspectives on activity and context*. Cambridge UP, 1996, pp. 1–32.

[17] M. T. H. Chi, "Quantifying qualitative analyses of verbal data: A practical guide," *The Journal of the Learning Sciences*, vol. 6, no. 3, pp. 271–315, 1997.

[18] R. Krauss and S. Fussell, "Constructing shared communicative environments," in *Perspectives on Socially Shared Cognition*, L. Resnick, J. Levine, and S. Teasley, Eds. American Psychological Association, 1991, pp. 172–200.

[19] E. Hutchins, "The technology of team navigation," in *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, J. Galegher, R. Kraut, and C. Egido, Eds. Lawrence Erlbaum Associates, 1990, pp. 191–220.

[20] C. Hmelo-Silver and M. Pfeffer, "Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions," *Cognitive Science*, vol. 28, pp. 127–138, 2004.

[21] M. Fowler and K. Scott, *UML Distilled: A brief guide to the standard object modeling language*, 2nd ed. Addison-Wesley, 2000.

[22] K. Severinson-Eklundh, H. Huttenrauch, and A. Green, "Social and collaborative aspects of interaction with a service robot." *Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots*, vol. 42, no. 3-4, 2003.

[23] 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*, vol. 19, no. 1–2, pp. 85–116, 2004.

[24] M. van Duuren and M. Scaife, ""Because a robot's brain hasn't got a brain, it just controls itself" - Children's attributions of brain related behaviour to intelligent artefacts," *European Journal of Psychology of Education*, vol. 11, no. 4, pp. 365–376, 1996.