Fostering Common Ground in Human-Robot Interaction*

Sara Kiesler

Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA 15232, USA
kiesler@cs.cmu.edu

***Abstract* – Effective communication between people and interactive robots will benefit if they have a common ground of understanding. I discuss how the common ground principle of least collective effort can be used to predict and design human robot interactions. Social cues lead people to create a mental model of a robot and estimates of its knowledge. People's mental model and knowledge estimate will, in turn, influence the effort they expend to communicate with the robot. People will explain their message in less detail to a knowledgeable robot with which they have more common ground. This process can be leveraged to design interactions that have an appropriate style of robot direction and that accommodate to differences among people.**

Index Terms* – *human-robot interaction, social robots, humanoids, perception, dialogue, common ground, knowledge estimation, speech communication

INTRODUCTION

Imagine a future situation in which robots at the Nashville USA airport perform as security guards and guides, directing people to their terminal and gate. How should these robots explain routes through the airport to locals and strangers, to young and old, to people in a hurry and to people with time to fill? The answers to these questions are not just a matter of whether the robot should use speech or should distribute printed maps—classic HCI questions. Since there will be little time for learning, the robots' appearance and initial behavior must create in visitors an appropriate mental model of the robots' abilities and intentions. The robots must be able to repair inevitable misunderstandings, and they must adapt to the needs of different travellers. For example, the robots may need to change their interaction style depending on travellers' age.

Thus far, we typically make these design choices using hunch, and trial and error. We might improve this approach by applying social psychological and cognitive theories to the human-robot interaction (HRI) design space. Theory can generate testable hypotheses about HRI design choices. Experiments to test these hypotheses could lead to advances in practice and in the science of human robot-interaction.

In this paper I explore the application of the theory of common ground to human robot interaction. The theory of common ground was developed to understand communication between people. Its main assumption is that communication between people requires coordination to reach mutual understanding, just as ballroom dancers and basketball teams do. The process of coordination relies on a large amount of shared knowledge between the parties, that is, common ground [1]. We take common ground for granted in many person-to-person interactions. For example, if you approach a human security guard at the Nashville airport, both the guard and you know the appropriate topics of conversation to initiate, who has authority and legitimacy for making different requests, and approximately who has knowledge about what. (You know where you want to go; the guard knows the security rules and where the exits are located.) Common ground makes it possible for you to approach the guard—a total stranger—and say, "which direction for U.S. Air?" and the guard will know how to answer, "That way" [points]."

In the following sections of this paper, I show how common ground applies to human-robot interaction in situations where robots interact with people in public spaces. In these situations, as in the Nashville airport, the robot, and the people who interact with it, usually are strangers. To form common ground, they must develop compatible mental models of one another. I discuss research demonstrating the tendency for people to attribute knowledge to a strange robot based on their beliefs about the robot's origin, the tendency for people to communicate with a strange robot differently based on the robot's physical characteristics, and the ways that a robot could speak to a person depending on the expertise level of the person. I call on these lines of evidence to support the argument that the common ground principle is an important factor in human-robot interactions.

COMMON GROUND WITH A ROBOT

One of the key postulates of the theory of common ground is least collaborative effort, that is, people in conversation minimize their collective effort to gain understanding [2]. For instance, the security guard might well gesture silently in the direction you are supposed to walk, and you, in turn, might say "thanks" to confirm you understand. Thus, in an efficient few seconds, with a single word, you and the guard assert and acknowledge common ground.

Achieving least collective effort should be an ultimate goal of successful human robot interaction. If you ask a robot guard for directions to "Gate 10," you do not want to listen to your help options or to a long recitation of directions to each gate in turn. You want a quick pointer to Gate 10. Many of today's dialogue systems lack common ground, that is, little understanding of callers' needs, and they cause no end of frustration. If eventually robots can conform to the principle of least collective effort, people will have to work less hard to communicate with them than they do today, and they are likely to feel the interaction is more satisfying.

* This work is supported by NSF ITR Grant #IIS-0121426.

Mental Model of a Robot

Ideally, a robot should not have to enumerate its functions and knowledge domains to people before they can interact with it. We want these functions and domains to be obvious. The solution for achieving this most basic form of common ground is to create in people's minds an appropriate mental model of the robot automatically. In particular, the robot should prompt people to make an appropriate estimate of the robot's role and what the robot knows.

Interactive robots have a start on this estimation process because they are, by virtue of their interactivity, somewhat humanlike. Interactivity in the form of speech or gesture, especially, will prompt observers to anthropomorphize automatically, without any intent or thoughtful processing (http://www.anthorpomorphism.org). Anthropomorphism also is likely when people see an animal or object that displays humanlike movements [3] or appears to act intentionally [4]. People are likely to assume these robots have humanlike roles and capabilities too.

Nass and his colleagues have argued that people apply stereotypes and social heuristics, and enact social habits with interactive systems automatically and mindlessly [e.g., 5, 6]. There is considerable evidence for automaticity in some aspects of social behavior [7]. However, automaticity does not preclude the influence of mental models. We argue that people who interact with a system create an implicit mental model of the system. The mental model reflects any anthropomorphism that has occurred, and it leads to expectations of the behavior of the system.

Consider an experiment we once ran in which subjects played a Prisoner's Dilemma game for real money with a real person or with a computer agent displayed on a screen [8]. (In this Prisoner's Dilemma game, if both partners cooperated and put up their money [$3.00 each], then they both gained $6.00. If one defected and kept his money, then that partner won $9.00 and the cooperator got nothing. If both defected, then they both got nothing.)

The results showed that when the agent looked like the person, people cooperated with the person-like agent at the same level as they did with the real person. When the agent acted exactly the same way, but looked like a dog instead of a person, on average, people's cooperation declined markedly. Yet, as shown in Figure 1, dog owners behaved differently than nonowners. The dog owners cooperated as much with the dog-like agent as with the person and person-like agent. Post-test questionnaire data suggested that the dog owners had more confidence that the dog-like agents would respond to their cooperative behavior with cooperation. More trust was associated with higher levels of cooperation.

These results suggest that the subjects in this experiment responded not just mindlessly to the system's appearance and behavior, but also to their mental model of what the system represented. We think dog owners and nonowners carried different expectations of what the dog agent would do. That is, they had different mental models of the dog agent, based not just on what they saw of the agent but also on their experience with dogs.

	Level of Cooperation with a Person and Person-like Agent		Level of Cooperation with Dog-like Agents	
	Dog Owners (n=18)	Non-Owners (n=30)	Dog Owners (n=16)	Non-Owners (n=32)
Round 1				
% of Subjects who Cooperated	89%	87%	81%	53%
All 6 Rounds*				
% Subjects who Cooperated	78%	69%	67%	52%

Fig. 1. Cooperation among dog owners and non-owners.
*In prisoner dilemma games, cooperation typically is higher on the early rounds and drops on the last round because the subjects can defect without the possibility of retaliation.

Why are these data relevant to common ground with robots? I believe they are relevant because they suggest that people's mental model of a robot will influence their expectations of the robot and therefore whether the robot is likely to achieve common ground and interact efficiently with them. In the experiment described above, whether or not subjects had a mental model of the dog agent as cooperative influenced whether the agent was able to form a cooperative agreement with them. Similarly, in interactions with a robot, people's mental model of a robot will influence whether or not the robot can reach mutual understanding with them.

The results of the experiment also suggest that mental models are not just general beliefs (e.g., this robot is nice, funny, or respectful). Mental models also comprise a set of task-specific expectations of process, that is, how the system will work. For example, a guard robot might be expected to know about airport gates and to point them out but it would not be expected to know locations in Nashville. In the Prisoner's Dilemma experiment, the subjects thought the dog-like agent was more attractive and charming than the person-like agent, but they cooperated with the person-like agent more than they did with the dog-like agent. This pattern happened because, even though they liked the dog-agent more, they expected the person-like agent to be more likely to understand their own cooperative strategy and to cooperate as a person would.

I also argue that mental models are situation specific, that is, that expectations of process can change depending on the situation. Thus, people might have similar expectations of a human guard and a robot guard in one task domain such as pointing out gates but different mental models of a human guard and robot guard in another task domain such as remembering retail landmarks at the airport (where is the closest Starbucks) [9, 10]. Further, people are likely to have different mental models of the same robot when it is enacting one task versus another, such as when it is being serious or playful [11].

How People Estimate a Robot's Knowledge

I have suggested that social cues, humanlike movement, and anthropomorphism, among other things, will influence people's mental model of a robot. We have begun to study one aspect of the mental models that people hold of robots, that is, their estimates of a robot's knowledge.

Knowledge estimation is an important part of grounding in communication. When we meet other people, we go through a knowledge estimation process in which we exchange information such as names, intentions, and so forth. To exchange information successfully, we estimate others' shared common knowledge and formulate our messages in respect to this shared knowledge [12]. For example, when a stranger asks you for directions to a local restaurant, you estimate or determine where the stranger lives. If you perceive that he lives in the local area, you also infer he knows the names of local landmarks, and you use these names to tell the person about the route to the restaurant. If you think the person is not local, you will not use the names of local landmarks in referring to the route.

Clark and his associates, e.g., [13], proposed that people use observable physical and linguistic cues, as well as information they have about each others' group memberships, educational background, or professional identities, to estimate each others' knowledge. People are highly accurate in their estimates of the distribution of mundane knowledge in a particular population. For example, students were able to estimate the proportion of other students who knew the names of public figures [14] and landmarks [15].

If people are unfamiliar with a robot, how can they make estimates of its knowledge? The previous work on social cues suggests that physical, linguistic, and social context cues will guide these estimates.

Sau-lai Lee and students in our lab conducted two controlled experiments to test the hypothesis that people's representation of a robot's knowledge would change when the origin of the robot changed [16]. Lee et al. proposed that a robot's origin, such as whether it is made in America or Asia, could be used as a cue to guide knowledge estimations. Thus, an American-made, English-speaking robot might be assumed to know better where the Empire State building is than a Hong Kong-made, Cantonese-speaking robot.

The subjects in the experiments were Chinese university students from Hong Kong who saw a video of a robot interacting with the experimenter. Half of the subjects were randomly assigned to a condition in which they saw the robot speak Cantonese with the experimenter (who was Chinese). They were told the robot was built at a robotics institute in Hong Kong. The rest of the subjects saw the robot speaking English with the experimenter, and they were told the robot was built at a robotics institute in New York.

Subjects in both conditions completed the knowledge estimation tasks. First they viewed a set of 14 Chinese and American landmarks. Next they were asked to view the landmarks one by one, and to identify the landmarks themselves. Finally, they were asked to estimate the likelihood using a rating scale from 0% likelihood to 100% likelihood that the robot could identify each landmark.

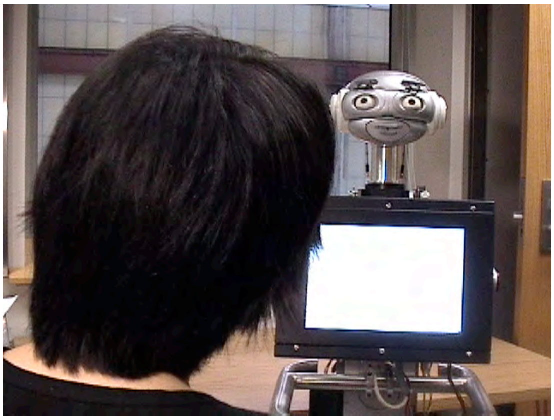

Fig. 2. Experimenter with robot, as seen by participants.

Lee et al. compared subjects' estimations of the robot's knowledge when the robot originated either in Hong Kong and spoke Cantonese or New York and spoke English. They hypothesized that the origin of the robot and language it used would create different mental models of the robot in the minds of subjects. Subjects should infer that both robots would have greater knowledge of famous landmarks than obscure landmarks. But also, subjects should believe the robot built in Hong Kong had knowledge of Hong Kong tourist landmarks, and that the robot built in New York had knowledge of New York tourist landmarks.

The results of this study, summarized in Figure 3, showed that subjects estimated the knowledge of the robot based in part on their assumptions about people, extrapolated to the robot. They expected the robot to know more of the landmarks that are famous in both countries (such as the Great Wall of China) and less likely to know the landmarks that are unfamiliar to people in both countries. Also, the origin of the robot influenced their estimations. The Chinese robot was perceived as more likely to know famous Hong Kong landmarks than the American robot. The American robot was perceived to know about New York landmarks only as well as the Chinese robot (an in-group bias).

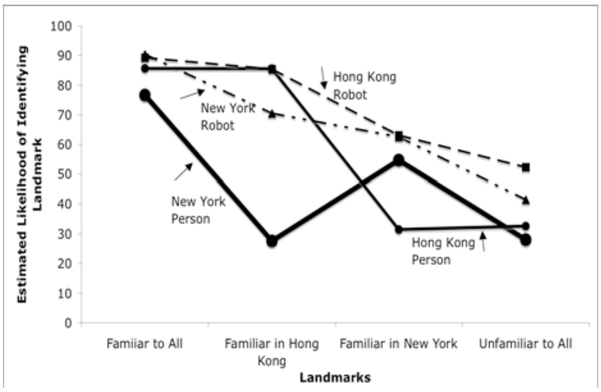

Fig. 3. Mean estimates of a person's or a robot's knowledge of landmarks. (Landmarks varied in their familiarity to residents of New York and Hong Kong.) The solid lines represent data from a human-human study [16]. The dashed lines are from a human-robot study [17].

Lee et al. also found the results to be highly correlated with the results of people's estimates of other people's knowledge, $r = .85$ in the HK condition and $r = .76$ in the NY condition. These correlations strongly suggest that subjects in the experiments used their knowledge of people as an anchor for estimating the robot's knowledge.

The results of this study suggest that given minimal information about a robot (languages it speaks; where it was created), people generalize to construct a rudimentary mental model of the robot's knowledge in a specific domain (tourist landmarks).

The data do not tell us how subjects justified these extrapolations. Did they believe that the Hong Kong (or New York) engineers who built the robot also put information about tourist landmarks into a database accessible to the robot? Did they believe the robot in Hong Kong (or New York) had direct experience with landmarks? Or did they believe that when the robot learned languages it also learned about names and places? Research suggests that any or all of these could be true. When considering other people and animals, we humans reflect on hidden causes of observed behavior, make attributions as to the traits, experiences, or reasons for this behavior, and extrapolate to new situations and competencies [18]. These tendencies are well established neurologically, and are likely triggered automatically by our observation of machines that have human attributes and move and speak purposefully, that is, by our anthropomorphism. If so, then knowledge estimates can co-exist with an assortment of post hoc meta-reasoning about these aspects of mental models. In other words, we may strongly believe, "this robot knows all about New York," with a few weak hypotheses about how the robot could have attained this state.

With or Without Common Ground, How People Talk with a Robot

Research has shown that people's estimates of others' knowledge significantly influences how they construct their communications. In one study, when subjects described public figures to another person, they provided descriptive information about the public figures in inverse proportion to their estimates that the other person could identify the public figure [14].

This work points to the strong possibility that when people interact with a robot, their estimates of the robot's knowledge will influence how they talk with the robot. For instance, if you need to send a robot to a location and you assume the robot is familiar with the terrain (that is, has common ground with you), your estimate of its knowledge should cause you to (a) use local landmarks to direct the robot, and (b) reduce the amount of information you give the robot because you assume the robot already knows the area.

To investigate this process, Aaron Powers and students in our lab conducted an experiment to study how a robot's persona and subjects' consequent estimates of the robot's knowledge would influence their assumption of common ground with the robot and their communication with it [19]. Subjects talked with the robot shown in Figure 2 through an interface like that of Instant Messaging (IM). The robot spoke aloud and displayed what it said on the IM display. The robot's persona was presented as either female (feminine voice, pink lips) or male (male voice, grey lips). Powers et al. used these two simple cues intentionally, to demonstrate that differences in robot persona could be accomplished through minimal variation of a robot's appearance and voice.

In the experiment, the robot was to be a dating counsellor in the future, and asked each subject about romantic dating practices, ostensibly to build its own store of knowledge. In human populations, women are far more knowledgeable about dating norms and social practices, and they have more social skills than men do [20]. Therefore, subjects should assume a "female" robot would know more about dating practices and norms than a "male" robot would, and, in reality, female subjects would know more about dating practices than male subjects.

According to common ground theory, subjects should describe and explain dating norms briefly (with fewer words and with linguistic shortcuts and jargon) to a female robot than to a male robot because the female robot would be assumed to share their dating knowledge. Further, women should assume more overlapping knowledge with a female robot than men, so should be particularly brief with a female robot. Finally, during the experiment, the robot asked the subjects about dating practices for a hypothetical couple, John and Jill. Since women in general have more dating knowledge than men, and female robots have more dating knowledge than male robots, the least elaborate communication should be found when women explain dating norms for women to a female robot.

The results of the study showed that both women and men, answering questions from the robot about dating, spoke most briefly with the female robot and at greatest length and detail with the male robot, especially about dating norms for John. Also, women said more to the female robot about John than about Jill, whereas men said more to the female robot about Jill than about John. The results support our thesis: people explain less to a robot they think already knows the subject matter, that is, to the robot with whom they share common ground.

DESIGNING HRI FOR COMMON GROUND

If our characterization of how people could interact with a robot is correct, then a straightforward design implication for the security guard robot at the Nashville airport is clear: Make the robot anthropomorphic, and make it look somewhat machinelike, mature, stern, and male [11]. Put it in uniform, or use a clear sign, to provide an instant assessment of its job. Airport visitors should be able to see that the robot's job is to guard, and that it probably knows where security lines are located in the airport. If we wanted this robot to have minimal and efficient conversation with visitors about airport security locations, then the robot should conform to the stereotype of guards. People will assume common ground with the robot (on location and security topics), and speak directly on point to the robot.

Suppose instead we wanted visitors to provide more detail, to speak slowly, to be redundant. Such a design goal might exist if the robot were not just a guard, but also functioned as a guide. Or perhaps we want people to elaborate to support a robot's poor speech understanding. Then we might want to consider a different interface, one that did not create such a strong impression of a security guard. For example, we could create a broader model in visitors of what the robot knew about airport locations by making it look more feminine and youthful, characteristics associated with jobs like docent, teacher, or nurse rather than security guard [11]. Since to achieve common ground people adapt their speech to the perceived needs of the other, they should adjust their speech to the needs of this robot. They are likely to speak more clearly and patiently to the more childish robot guard.

The Robot's Mental Model

The theory of common ground can provide guidance for a mixed initiative design approach in which a robot takes some responsibility for achieving common ground with people. Indeed, the mixed initiative approach is consistent with the common ground assumption that both partners in an exchange must cooperate to achieve common ground. For a robot to help achieve common ground in public places, it would need to detect people and differences among people. For instance, the robot might detect who, in a group of pedestrians, is attending to the robot and wants help. It might detect whether the person is in a wheelchair, is carrying luggage, is headed for the gates, and other information that would be helpful in constructing an appropriate message. The more the robot can adapt to the person, the less the person needs to adapt to the robot.

Accommodating to Different People

Cristen Torrey, a graduate student in our lab, has developed a compelling argument for designing robots that adapt their messages appropriately to the needs of their users [21]. By doing so, robots are more likely to achieve the goal of least collective effort.

Torrey reviewed the literature on "elderspeak"—a tendency for people to talk with older people in institutions as though they were toddlers. Nurses, doctors, visitors, and others use simpler vocabulary with old people, minimize their number of words, and repeat themselves. They also exaggerate their intonation, speak in a high pitched voice, speak slowly, and make a number of controlling gestures such as patting the person on the head [22]. Torrey concluded that some forms of elderspeak, for instance, stressing the important words in a sentence, can help elders to understand complex language whereas other forms of elderspeak, such as terms of endearment, do not help. Furthermore, elderspeak, even if it aids communication, can make listeners feel less competent and disrespected.

Torrey is experimentally testing the design implications of this argument for a robot's communications with people whose needs for information vary. She argues that when people have expertise in a domain, a robot that provides too much clarification or detail about information in that domain will constitute "overkill" and will violate the least collaborative effort principle. Too much elaboration will seem to be disrespectful of the person's skill. Too little detail would have the same impact on a novice. Thus a novice will welcome the same clarification and detail that is unsuitable for an expert. To achieve common ground, then, a robot should adapt to the person's level of knowledge and the degree of common ground between the person and the robot. Thus, our robot guard should assess whether a visitor to the airport is familiar with the airport (a Nashville resident) or a stranger. The resident is more likely to know where gates are by reference to landmarks whereas the stranger is not. By adjusting to the likely knowledge of the resident or stranger, the robot can reach common ground with least collaborative effort.

Conclusion

Scassellati has argued that a humanoid robot needs to have a theory of mind [23]. We take this argument a step further to argue that an interactive robot should have a theory of common ground and should incorporate action plans that will achieve common ground with people. Social cues emitted by the robot are a first step in creating common ground, as these cues can elicit in people an appropriate mental model of the robot. The robot's initial behaviors can correct people's mental models if they are inappropriate and they can repair damage to common ground. Beyond these initial steps, the robot will need to assess the expertise and needs of individuals, so that its actions reflect the requirements for common ground across different people.

Acknowledgements

I am grateful to Susan Fussell, Cristen Torrey, Aaron Powers, and an anonymous reviewer for their helpful suggestions for this paper.

References

[1] H. H. Clark, "*Arenas of language use*," Chicago: University of Chicago Press, 1992.

[2] Clark, H. H., & S. E. Brennan, "Grounding in communication." In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA, 1991.

[3] Morewedge, C.K., Preston, J., & Wegner, D.M. "Timescale bias: Anthropotempocentrism in the ascription of mental states." Talk. Society for Personality and Social Psychology, New Orleans, January 2005. (Obtain ms. from morewedg@wjh.harvard.edu). Scholl, B. J., & Tremoulet, P. D. "Perceptual causality and animacy." *Trends in cognitive science*, Vol. 4, pp. 200-309, 2000.

[4] C. Nass, and Y. Moon. "Machines and mindlessness," *Journal of Social Issues*, vol. 56, no. 1, pp. 81-103, 2000.

[5] C. Nass, Y. Moon, and P Carney, "Are respondents polite to computers? Social desirability and direct responses to computers," *Journal of Applied Social Psychology*, vol. 29, no. 5, pp. 1093-1110, 1999.

[6] J. A. Bargh, M. Chen, and L. Burrows, "Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action," *Journal of Personality and Social Psychology*, vol. 71, no. 2, pp. 230-244, 1996.

[7] S. Parise, S. Kiesler, L. Sproull, and K. Waters. Cooperating with life-like interface agents. *Computers in Human Behavior*, vol. 15, pp. 123-142, 1999.

[8] N. Shechtman and L. M. Horowitz "Media inequality in conversation: How people behave differently when interacting with computers and people," *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems*, New York, NY, 2003.

[9] Morkes, J., H. Kernal, and C. Nass. "Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of social responses to communication." *Human-Computer Interaction*, vol. 14, no. 4, pp. 395-435, 2000.

[10] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation." Robot and Human Interactive Communication, 2003. Proceedings of ROMAN 2003 (pp. 55-60). The 12th IEEE International Workshop on Robot and Interactive Communication, Vol. IXX, Oct. 31-Nov. 2, 2003, Milbrae, CA.

[11] R. S. Nickerson, "How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others," *Psychological Bulletin*, vol. 125, no. 6, pp. 737-759, 1999.

[12] E. A. Issacs and H. H. Clark, "References in conversation between experts and novices," *Journal of Experimental Psychology:General*, vol. 116, no. 1, pp. 26-37, 1987.

[13]S. Fussell and R. Krauss, "Coordination of knowledge in communication: Effects of speakers' assumptions about what others know." *Journal of Personality and Social Psychology*, vol. 62, pp. 378-391, 1992.

[14]I. Y-M. Lau, C. Chiu, and Y. Hong, Y. "I know what you know: Assumptions about others' knowledge and their effects on message construction." *Social Cognition*, vol.19, pp. 587-600, 2001.

[15]S-L. Lee, S. Kiesler, I.Y. Lau, and C.Y. Chiu, "Human Mental Models of Humanoid Robots," *2005 IEEE International Conference on Robotics and Automation*, April 2005.

[16]S-L. Lee & C. Chiu, "Communication and shared representation: The role of knowledge estimation." Unpublished. Hong Kong University, HK., 2003.

[17]D. J. Povinelli and J. M. Bering, "The mentality of apes revisited," *Current Directions in Psychological Science*, vol. 11, no. 4, pp. 115-119, August 2002.

[18]A. Powers, A. Kramer, S. Lim, J. Kuo, S-L. Lee, S. Kiesler, "Common Ground in Dialogue with a Gendered Humanoid Robot," Proceedings of ROMAN 2005. The 14th IEEE International Workshop on Robot and Interactive Communication, Vol. IIIXX, Nashville, TN. Aug 13-15, 2005.

[19]W. Wood, N. Rhodes, M. Whelan, M., "Sex differences in positive well-being: A consideration of emotional style and marital status." *Psychological Bulletin*, vol. 106, pp. 249-264, 1989.

[20]C. Torrey, "Appropriate accommodations: Speech technologies and the needs of older adults." Unpublished ms. HCII, Carnegie Mellon University, May, 2005.

[21]E. B. Ryan, M. L. Hummert, and L. H. Boich. (1995). Communication predicaments of aging: Patronizing behavior toward older adults. *Journal of Language and Social Psychology*, vol. 14, no.1-2, pp. 144-166, 1995.

[22]B. Scassellati, "Theory of mind for a humanoid robot," Autonomous Robots, vol. 12, pp. 13-24, 2002.