To appear in E. Morsella (Ed.), Expressing oneself/Expressing One's Self: A Festschrift in Honor of Robert M. Krauss. Taylor and Francis. In Press.

Running head: ROBOTS AND PERSPECTIVE-TAKING

What Robots Could Teach Us about Perspective-Taking

Cristen Torrey, Susan R. Fussell and Sara Kiesler

Human-Computer Interaction Institute

Carnegie Mellon University

04/01/07

Correspondence:

Cristen Torrey

Human-Computer Interaction Institute

Carnegie Mellon University

5000 Forbes Avenue

Pittsburgh, PA 15213

412 268-4059

ctorrey@cs.cmu.edu

What Robots Could Teach Us about Perspective-Taking

Cristen Torrey, Susan R. Fussell and Sara Kiesler

Carnegie Mellon University

A distinctly social ability that underlies shared meaning, empathy, and cooperation is taking the perspective of another person during conversation. Research on communication has explored the manner in which human speakers account for their listeners' perspectives and adjust their communications in their attempts to be understood (e.g. Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Krauss, Vivekananthan & Weinheimer, 1969). Speakers attend to their listeners' group memberships and likely areas of expertise as they construct their messages (e.g., Clark & Marshall, 1981, Fussell & Krauss, 1992; Hupet, Chantraine & Neff, 1993; Isaacs & Clark, 1987). Speakers attend to what their partners can see, that is, their spatial perspective within the environment (e.g., Gergle, Kraut & Fussell, 2004; Lockridge & Brennan, 2002; Kraut, Miller & Siegel, 1996; Schober, 1993). And speakers attend to the verbal and nonverbal responses of their listeners to assess whether their message is comprehended and to make appropriate repairs and adjustments (e.g., Clark and Wilkes-Gibbs, 1986; Krauss & Bricker, 1966; Krauss & Weinheimer, 1966; 1968). These adjustments produce more effective communication, whether in the context of a single message (e.g., Fussell & Krauss, 1989) or over the course of an ongoing conversation (Kraut, Lewis & Swezey, 1982; Schober & Clark, 1989).

There now exists considerable evidence on the information people use in perspective-taking (for a recent review, see Schober & Brennan, 2003), but we know very little about failures in perspective-taking and how conversationalists cope with inaccurate or inadequate perspective-

taking. We suggest that the specific nature of perspective-taking in effective communication becomes particularly visible in conversations between humans and machines. Currently, the most common experience people have conversing with a machine is when they make a phone call to a customer service department and are greeted by an automated representative. These computer-driven speakers communicate with limited, if any, perspective-taking abilities. These speakers have no sense for the caller's familiarity with the task in question or with the number of times the caller has already been forced to listen to the complete set of instructions. Nearly everyone has a frustrating story to tell about these automated helpers. Many of these frustrations can be traced to an absence of perspective-taking skill on the part of the machine. These automated conversational partners do not have any sense for their listeners.

It seems likely that our conversations with machines might benefit from their use of perspective-taking strategies. Yet it is not at all clear how automated help systems, computer agents, or robots will be able to take the individual perspectives of their listeners. In ongoing research, we are exploring perspective-taking between humans and automated conversational partners, particularly in the embodied form of humanoid robots. Because machines are literal, they must be told precisely how perspective-taking should unfold. This necessity highlights some gaps in our understanding of perspective-taking, especially when it must be created or repaired.

When attempting to implement perspective-taking theories in a robotic form, we need to elaborate the details of the theory in a precise, computational manner. For example, if the robot predicts that the listener does not know a referent, we must specify the number of additional words of description that the robot should add. Thus far, theory informed by previous research underspecifies the answer to this and similarly specific questions about the process of perspective-taking. Because of these challenges, we believe that conversational robots offer a

unique opportunity to investigate the role of perspective-taking in effective conversation in a controlled manner and to observe the consequences of poor perspective-taking. In this chapter, we begin by describing empirical studies exploring the presence and consequences of perspective-taking in human-robot communication. We then describe our attempts to implement perspective-taking strategies with robotic conversational partners. We conclude by posing some issues still to be explored in understanding conversational perspective-taking behavior.

Robots as Conversational Partners

Research on humanoid robotic machines has made impressive progress. Part of the motivation behind the development of humanoid robots is the ease with which people relate to machines socially when these machines give anthropomorphic cues such as humanlike form (Powers & Kiesler, 2006) and speech (Nass & Lee, 2001). Many roboticists argue that robots with human form and language will be more effective communicators than machine-like robots because they will evoke familiar social responses in the humans with whom they interact. If these robots can live up to the expectations their social forms create, then getting information from them and working alongside them should be easier as well (Scassellati, 2004).

We are pursuing perspective-taking as a feature in the development of intelligent human-robot communication. We believe perspective-taking may be particularly important for robots interacting with a varied group of individuals in the role of an advisor, instructor, or guide. Robots in roles such as these may give tours in a museum, guide people in airports or shopping malls, tutor students, or answer questions in an information kiosk. As they interact with people of different backgrounds and levels of expertise, it may be advantageous for such robots to have the capacity to adjust their communication using perspective-taking strategies. Our initial approach to understanding perspective-taking between humans and robots begins by asking two questions

designed to understand the appropriateness of the perspective-taking approach for human-robot communication. First, do human speakers assume the robot has a perspective? And second, if robots were able to take their human listeners' perspective, would listeners benefit?

Do Speakers Take a Robot's Perspective?

We have found that people do make assumptions about the knowledge a robot has and does not have, using the same sorts of cues that people use when making assumptions about one another's perspectives. In one study, participants were asked to estimate the likelihood that a robot would recognize different landmarks (Lee, Kiesler, Lau & Chiu, 2005). The landmarks shown were familiar to residents of Hong Kong, familiar to residents of New York, familiar to both, or familiar to neither. When the robot was introduced as a research project of a New York university, participants estimated that the robot was more likely to recognize New York landmarks. When the robot was introduced as a research project of a Hong Kong university, participants estimated that the robot was more likely to know Hong Kong landmarks. This experiment suggested that people estimate a robot's knowledge differently based on the robot's "nationality," in much the same way as they do with other people (Fussell & Krauss, 1991; Isaacs & Clark, 1987).

The landmarks experiment tested people's predictions about what the robot was likely to know based on where the robot was built. If people predict that the robot is unlikely to know something, we would expect that prediction to result in more descriptive messages (Fussell & Krauss, 1992). In a new experiment, we tested whether participants would make assumptions about a robot's "gender" from its voice and appearance, and tested whether these assumptions would translate into changes in participants' communicative behavior (Powers, Kramer, Lim, Kuo, Lee & Kiesler, 2005). Participants were asked to instruct a humanoid robot in the modern

rules of dating. The gender of the robot was manipulated using the color of the robot's plastic lips (red or grey) and its voice pitch (higher or lower). Participants were told that the robot was gathering information to become a dating counselor. The robot asked participants a series of questions about what typically happens on dates. For example, the robot asked participants how to set up a date, who should do the planning for the date, and whether either member of the couple should buy new clothes for the date. Overall, participants used a greater number of words when describing dating to a male robot, suggesting that the male robot was perceived to have less knowledge of dating norms than the female robot. Further, male participants said more to the female robot than the male robot while female participants said more to the male robot than the female robot. We suggest that participants were using their own knowledge as a guide in predicting what the robot knew (Fussell & Krauss, 1992; Nickerson, 1999). When the participants' and the robot's gender overlapped, the participants may have perceived that there was less need for descriptive detail. In this study, we found that participants constructed different messages depending on the robot's gender (and the stereotypes that go with it), as well as their own similarity to the robot.

----- Insert Figures 1 and 2 About Here -----

Do Listeners Benefit from a Perspective-Taking Robot?

The previous work provided some evidence that speakers were considering the robot's perspective and adjusting their messages accordingly. In subsequent research, we considered whether there were advantages to having robots take their listeners' perspectives and adjust messages to listeners' expertise (Torrey, Powers, Marge, Fussell & Kiesler, 2006). In human-human communication, messages designed specifically for a listener are understood more easily than messages created for someone else or a generic listener (Fussell & Krauss, 1989; Krauss et

al., 1968; Kraut et al., 1982; Schober & Clark, 1989). In this study, we explored the benefits and consequences of adaptive communication on conversational efficiency, and we used a post-conversation questionnaire to investigate listeners' perceptions of the robot, the task, and the conversation. We were particularly interested in the extent to which appropriate perspective-taking behavior improved social relations with the robot. Prior research on the maintenance of "face" in conversation suggests that speakers may insult their listeners by ignoring their needs (e.g. Goffman, 1955; Holtgraves, 2002), but we are not aware of any empirical communication research that has tested the impact of the appropriateness of a speaker's perspective-taking communication on the listener's impressions.

We asked participants in a laboratory study to find and select ten cooking tools from sets of pictures on a computer monitor. A robot directed them to find each tool in turn and responded to any questions participants had about the tool, its size, for example, or its shape. Participants who signed up for the study were pre-tested for cooking expertise by completing a short quiz on cooking methods. We used their knowledge of cooking methods to inform the robot's behavior because pre-testing showed that this knowledge is highly correlated with people's knowledge of cooking tools. Thus, the robot could use the information that the participant knew how to sauté, for example, to infer that the participant also had some knowledge about whisks and silicone spatulas. We selected participants so that half were "experts," meaning they got a perfect score on the pre-test, and half were "novices," meaning they scored less than fifty percent correct on the pre-test. These participants interacted either with a robot whose perspective-taking communication was designed for experts ("Now, we need a paring knife") or a robot whose perspective-taking communication was designed for novices ("Now, we need a paring knife. It's usually the smallest knife in the set. It has a short, pointed blade that is smooth, not jagged.").

Participants were told they could ask the robot for help. We measured the number of questions participants asked the robot, their task performance, and used questionnaire measures to investigate participants' perceptions of the robot, the task, and their communication with the robot.

----- Insert Figure 3 About Here -----

Our results showed that novice users were affected disproportionately by a lack of information in the robot's directions. When the robot introduced the tool by its proper name alone, novices asked twice as many clarifying questions as experts. These questions extended the amount of time novices spent completing the task, but surprisingly it did not influence any of the questionnaire measures. The extra time novices spent on the task and experts' unneeded interaction with the robot did not negatively affect their perception of the robot as a conversational partner, as we had supposed it might. We surmised that participants did not care if their task performance was inefficient, and that participants might have even enjoyed conversing with the robot. Therefore, in a follow-up experiment, we offered a small monetary bonus to participants if they managed to complete the task quickly. When participants were particularly motivated to work quickly, the robot offering unnecessary description was rated as less effective, less authoritative, and more patronizing. Expert participants found the robot to be more effective, more authoritative, and less patronizing when the robot used only names of tools as a guide rather than extra description. These experiments suggest that while all participants benefited from an appropriate level of detail in the robot's communication, it was only participants with a particular demand on their time who evaluated the robot negatively if it did not have good perspective-taking abilities.

----- Insert Figure 4 About Here -----

Building a Perspective-Taking Robot

Our results suggest that people assume a robot has a perspective, and they adjust for it. Participants in the cooking tool study, however, did not seem to have a strong expectation that the robot should adjust to their perspective, even though they performed better when their level of expertise was accommodated. The idea of "least collaborative effort" has been proposed to describe the joint endeavor that human speakers engage in when they communicate (Clark & Wilkes-Gibbs, 1986). Although effort expended in a conversation need not be distributed perfectly, we generally assume that both parties in a conversation share the effort to create joint meaning. In a conversation between peers, both parties are making adjustments to communicate effectively, efficiently, and respectfully. When people and robots communicate, the appropriate distribution of effort is not as clear. One could argue that humans, being more flexible than computers, should bear responsibility for adjusting their communication to be understood. On the other hand, one could argue that robots are built to assist in the achievement of human goals, and their design should minimize human effort. Under this assumption, if robots were able to read human minds, so much the better. Our previous work demonstrates that people seem to have a fairly automatic tendency to take the robot's perspective, and yet they do not expect perspective-taking from a robot in a reciprocal way. Based on our participants' improved performance when the robot was using a form of perspective-taking, we believe it may be worth the effort to develop perspective-taking strategies that can be used by robots. This is particularly true when robots are providing instruction, directions or other types of information and may interact with people of varying levels of expertise.

But how can perspective-taking strategies be practically implemented on a robot? One possibility is a user modeling approach. By user modeling, here, we mean that the robot has a

model, or knowledge, of an individual's expertise or attitude. This approach requires that the robot have probabilistic knowledge of how expertise is distributed in the population. Then, when interacting with an individual, the robot, based on initial information, could make assumptions about the listener's position in the distribution and what the listener is likely to know. The pre-test we used in the cooking tools experiment is a crude example of this approach. By pre-testing participants, the robot gathered information about an individual's level of expertise; the robot could then use these assumptions to plan its communication about further tools.

But extensive planning of utterances to address listeners' perspectives may not be necessary. An alternative (or supplementary) approach involves the robot offering a small amount of information, for example, the proper name of a tool and then watching carefully for signals that the name is accepted by the listener. For instance, the listener may provide a backchannel utterance like "mm-hmm" or "uh-huh" that signals acceptance. The robot might also attend to task activity, so that if the listener's expected action were not taken in a timely manner, the robot could automatically initiate a repair. These two approaches are by no means the only ways of building perspective-taking abilities into a robot, nor are the two approaches mutually exclusive. There is no reason why a robot might not use both approaches simultaneously. In the sections that follow, we discuss how these two approaches to perspective-taking might be developed on a conversational robot.

Adjusting a Probabilistic Model

In our previous experiment with a cooking tool selection task, we inferred our participants' knowledge of cooking tools by quizzing them about several cooking methods. By gauging their knowledge of cooking methods, we could infer their likely expertise on cooking tools. People who know how to poach an egg are also likely to know the names of quite a

number of cooking tools. This general approach could be expanded into a more complete user model for use by a robot, specific to the cooking domain. When a person demonstrated knowledge of the word "poach" while conversing with the robot, the robot would calculate the likelihood that the person has other cooking knowledge based on the distribution of such knowledge in the population. In that case, the robot could be fairly confident that the person also knew how to sauté, for example, and it would not need to elaborate on such a direction. A user model, as just described, would require a model of how domain knowledge is distributed in the population, most likely obtained through surveys. This distribution would need to be created for different domains and for different groups of people with whom the robot might interact.

To be feasible, this general approach would require the specification of numerous details. With what sort of model does the robot begin? Is there an efficient order when introducing information, such that the robot models the listener in the quickest possible way? Does the robot adjust its user model based on performance cues only? Or should the robot also adjust based on affective cues such as a frustrated inflection in the listeners' voice? What are the specific features that make the robot's communication appropriate for listeners at different levels of expertise? There are no specific guidelines we can draw from theories of human-human communication, but future work with robots offers a unique opportunity to investigate the application of these research questions. With robots, questions about perspective-taking can be investigated in controlled ways where the robot interacts in precisely the same way with each participant.

Reacting to Grounding Cues

Even when they have little knowledge of others, speakers can adjust to the requirements of their listeners by paying close attention to the effect of their communication (Clark & Wilkes-Gibbs, 1986). Rather than expending effort up front in constructing the precisely appropriate

utterance for a listener, a speaker may make a reasonable attempt. Speakers need not wait for the listener to make explicit requests to make a repair; in fact, they seem to prefer to initiate the repair themselves (Sacks, Schegloff & Jefferson, 1974). There are numerous cues that speakers might use to confirm that their utterance is accepted or as evidence that a repair is necessary. Speakers attend to their listeners' verbal responses, including backchannel communications or lack thereof. If a listener uses an "uh-uh" or "ok" to confirm each step of a direction, when that backchannel communication is absent, the speaker may attempt a repair (Gergle, Kraut & Fussell, 2004b). If the speaker can see the listener's activities, the speaker can watch to see if the listener makes the expected movements and repair if those movements are not made (Brennan, 2004, Gergle et al., 2004b). By attending to these verbal and nonverbal communicative elements, speakers can initiate repairs before listeners have to ask questions or make explicit requests for a repair.

We attempted this approach to perspective-taking behavior in another experiment using the cooking tool selection task explained previously. We implemented two ways the robot could have awareness of the listeners' activity, that is, through gaze awareness and task activity awareness. The manipulation of gaze awareness we used in this experiment made use of an eye contact sensor that could roughly indicate whether or not the participant was looking at the computer monitor on which the task was displayed. Our model of gaze behavior followed an empirical model proposed by Nakano, Reinstein, Stocky, and Cassell (2003). In their study, Nakano et al. observed that speakers attended to their listeners' gaze when a new referent was introduced. If the listener's gaze moved to the referred object, then that object was grounded in the conversation. But if the listener continued to gaze at the speaker, the speaker understood that elaboration was required. In the context of the cooking tool selection task, the robot assumed

participants were working on the task and needed no help when they were looking at the monitor that displayed the pictures of the cooking tools. When participants were not looking at the monitor, the robot assumed they were looking back at the robot to ask a question or to re-read the directions written on the screen. When the robot became aware that the participant was not attending to the monitor, the robot offered an additional unit of information to help the participant make his or her selection. For example, the robot asked the participant to select the paring knife, and if the participant looked back at the robot without selecting a tool the robot said, "The blade is smooth, not jagged." In addition to gaze awareness, we also manipulated task activity awareness. From pre-testing, we knew that participants who knew the correct cooking tool could find and select it within four seconds. We therefore gave the robot a simple timer, set to four seconds, such that if, after being directed to choose a tool, the participant had not made a selection in that amount of time, the robot offered an additional unit of information. This approach assumes that when participants have not made a selection in a given time period, they do not recognize the name of the tool and require further elaboration.

----- Insert Figure 5 About Here -----

We explored these two types of awareness in preliminary experimental treatments that contrasted a robot's use of gaze awareness, both gaze and task activity awareness, and neither form of awareness (Torrey, Powers, Fussell & Kiesler, 2007). Participants interacting with a robot with both forms of awareness asked fewer questions than participants interacting with a robot with neither form of awareness. However, these awareness strategies did not improve participants' task accuracy or their time on task. Participants made the same number of mistakes regardless. Clearly, a robot's simply having a perspective-taking strategy is insufficient to improve shared meaning. One question, in particular, is whether the specific help given by the

robot to participants was appropriate to their needs. The robot had a list of additional information, to be given out one piece at a time, but these elaborations were not associated with the mistake the participant had made previously or, for example, where on the screen the participant's mouse had been hovering. It may not be enough to consider only that the listener is not taking action and needs more information. An important aspect of a truly intelligent, perspective-taking robot would be the ability to choose the specific kind of help that is necessary for each individual at a particular point in time. (Recall your experiences with an automated telephone help systems. The most sophisticated of these can understand that you need more help but they frequently offer the wrong kind of help.) A perspective-taking robot may need to account not only for when a listener needs help, but what specific bit of information the listener needs, and, further, how that information should be phrased.

Conclusion

Our research on perspective-taking in human-robot interaction shows many similarities between this process and perspective-taking in interpersonal communication. People make assumptions about what robot partners know, based on their attributes, and these assumptions guide how they formulate their messages. People are also sensitive to how well their robotic partners' take their own perspective into account. On the whole, the field of human-robot interaction has benefited from the vast body of prior research on human perspective-taking.

While attempting to develop a perspective-taking robot from existing theory, however, we have encountered a number of aspects of perspective-taking that are underspecified in our current theories of human perspective-taking behavior. What are the cues a robot might best use to make assumptions about a listener's perspective? Group membership and its related expertise have interesting effects on the way speakers produce messages. But how do speakers recognize

expertise by virtue of group membership in their listeners? Is speakers' recognition of expertise an all-or-nothing decision (in the way we have implemented it)? Given the speed and naturalness of conversation, it seems likely that many interactions begin that way. And once the speaker has decided upon the listener's expertise, how, precisely, does he or she provide the requisite level of detail in a message? Should cooking tools, for example, be described for less knowledgeable listeners in terms of shape, size, color, usage, or other features? How much information should be provided in each utterance, before pausing for feedback from the addressee?

Despite the challenges in developing perspective-taking behavior at the level of detail required by a computer program, the use of a conversational robot in testing these decisions is a unique opportunity. When we investigate these issues with a conversational robot, the robot's behavior can be strictly controlled. For example, it is possible to create a robot that provides specific types of information to addressees, or does or does not follow conventions of eye gaze, interruption, elaboration, and repair. It is an interesting opportunity to test features of communication that humans are not likely to do on command in the laboratory, such as ignore the listeners' perspective. We know that errors in perspective-taking do occur, and clever techniques using confederate speakers have been developed to study such errors on the addressees' side (e.g., Keysar, Barr, Balin & Brauner, 2000). With robotic speakers, we can more thoroughly investigate the communicative and affective consequences of various kinds of perspective-taking strategies and errors, thereby contributing to our theoretical understanding of the mechanisms of the perspective-taking process.

The consequences of inadequate or inaccurate perspective-taking assumptions is an important question, both practically and theoretically. For the foreseeable future, robots are unlikely to be perfect conversational partners, and it is important to understand how poor

perspective-taking abilities might affect human performance and impressions of robots. Affective reactions to poor perspective-taking are particularly intriguing and under-studied in the prior human communication literature. For example, what degree of error (e.g., talking to a full professor as if he or she were an undergraduate vs. an assistant professor), and how many errors are necessary before the listener reacts emotionally, for example, by feeling insulted, disliking the speaker, or getting angry? With robot speakers, we can investigate these issues by manipulating communication while keeping the speakers' other characteristics carefully controlled.

In conclusion, we have attempted to show in this chapter how the fields of human communication and human-robot interaction can shape and inform one another in such a way that both fields are advanced. The human research on perspective-taking has fostered productive research on how we should design humanoid robots to converse in social settings; at the same time, the research on robot perspective-taking has generated important questions for basic theory and in addition offers special opportunities for answering these questions.

Acknowledgments

The authors are grateful for assistance provided by our many collaborators over the years: Sau-lai Lee, Ivy Yee-man Lau, Chi-Yue Chiu, Aaron Powers, Adam Kramer, Shirlene Lim, Jean Kuo, Matthew Marge, Hau-Yu Wong, and the members of the Project on People and Robots. This research was supported by National Science Foundation ITR #IIS-0121426, HSD #IIS-0624275, and IGERT #DGE-0333420.

References

Brennan, S.E. (2004). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *World Situated Language Use: Psycholinguistic,*

Linguistic and Computational Perspectives on Bridging the Product and Action Traditions. Cambridge, MA: MIT Press.

Clark. H.H. & Marshall, C.E. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber and I.A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

Fussell, S.R. & Krauss, R.M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 21, 445-454.

Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology, 21*, 445-454.

Fussell, S.R. & Krauss, R.M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62, 378-391.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004a). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology. 23,* 491-517.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004b). Action as language in a shared visual space. *Proceedings of the Annual Conference on Computer-Supported Cooperative Work (CSCW 2004)* (pp. 487-496). NY: ACM Press.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, 19, pp. 213-231.

Holtgraves, T. Face management and politeness. in Holtgraves, T. (ed.) *Language as Social Action: Social Psychology and Language*, Lawrence Erlbaum, Mahwah, NJ, pp. 37-63.

Hupet, M., Chantraine, Y., & Neff, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8,* 339-346.

Keysar, B., Barr, D.J., Balin, J.A; Brauner, J.S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11,* 32-38.

Krauss, R. M., & Bricker, P. D. (1966). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America, 41*, 286-292.

Krauss, R.M. & Fussell, S.R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, 9, 2-24.

Krauss, R.M., Vivekananthan, P. S., & Weinheimer, S., (1968). "Inner speech" and "external speech": Characteristics and communication effectiveness of socially and nonsocially encoded messages. *Journal of Personality and Social Psychology, 9*, 295-300.

Krauss, R. M., & Weinheimer, S. (1964). Changes in the length of reference phrases as a function of social interaction : A preliminary study. *Psychonomic Science, 1,* 113-114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4,* 343-346.

Kraut, R. E., Lewis, S.H., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology, 43,* 718-731.

Kraut, R. E., Miller, M. D., & Siegel, J. (1996) Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proceedings of CSCW'96* (57-66). NY: ACM Press.

Isaacs, E.A. & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26-37.

Lee, S., Kiesler, S., Lau, I.Y. and Chiu, C. (2005). Human mental models of humanoid robots. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA '05)*. Barcelona, April 18-22., 2767-2772.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, 9, 550-557.

Nakano, Y., Reinstein, G., Stocky, T. & Cassell, J. (2003) Toward a Model of Face-to-Face Grounding. *Proceedings of 41st Annual Meeting of the Association for Computational Linguistics*.

Nass, C. & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*, 171-181.

Nickerson, R.S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737-759.

Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes, 218 - 225. *Conference on Human-Robot Interaction 2006*. Salt Lake City, March 1-3.

Powers, A., Kramer, A., Lim, S., Kuo, J., Lee, S. & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. *Proceedings of the 2005 IEEE*

International Workshop on Robots and Human Interative Communication (RO-MAN 05). Nashville, TN, USA, Aug. 2005, 158-163.

Sacks, H., Schegloff, E. & Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, vol. 50, no. 4, 696-735.

Scassellati, B. (2004). Theory of mind for a humanoid robot. *Autonomous Robots*, 12, 13-24.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1-24.

Schober, M. F. & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In Graesser, A., Gernsbacher, M., and Goldman, S. (Ed.) *The Handbook of Discourse Processes*, (pp. 123-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Schober, M.F. & Clark, H.H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.

Torrey, C., Powers, A., Marge, M., Fussell, S.R. and Kiesler, S. (2006). Effects of Adaptive Robot Dialogue on Information Exchange and Social Relations. *Proceedings of the Conference on Human Robot Interaction* (HRI ’06). ACM Press.

Torrey, C., Powers, A., Fussell, S.R. and Kiesler, S. (2007). Exploring Adaptive Dialogue Based on a Robot’s Awareness of Human Gaze and Task Progress. *Proceedings of the Conference on Human Robot Interaction* (HRI ’07). ACM Press.

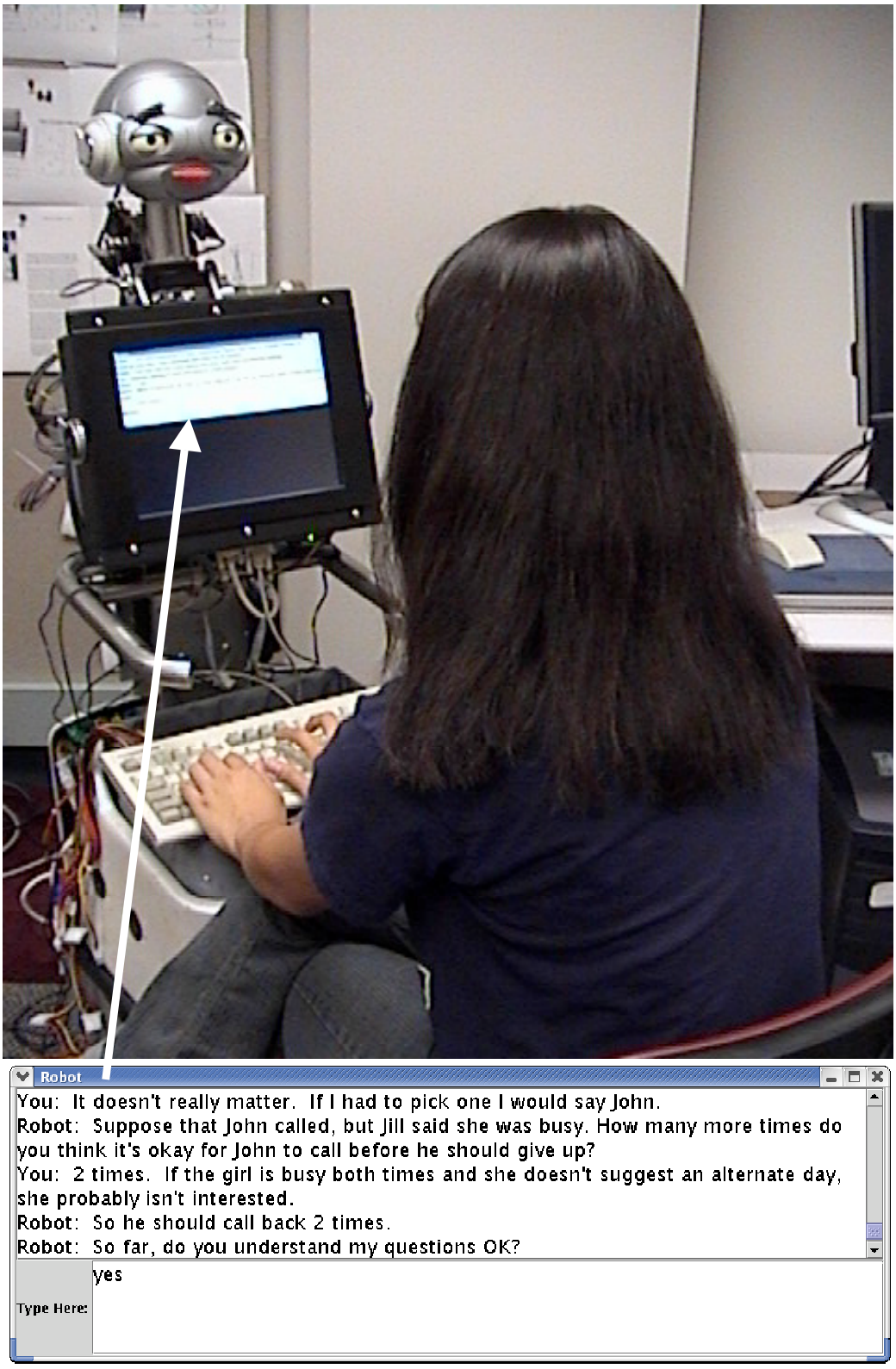

Figure 1. Interacting with a talking robot about dating (adapted from Powers et al., 2005).

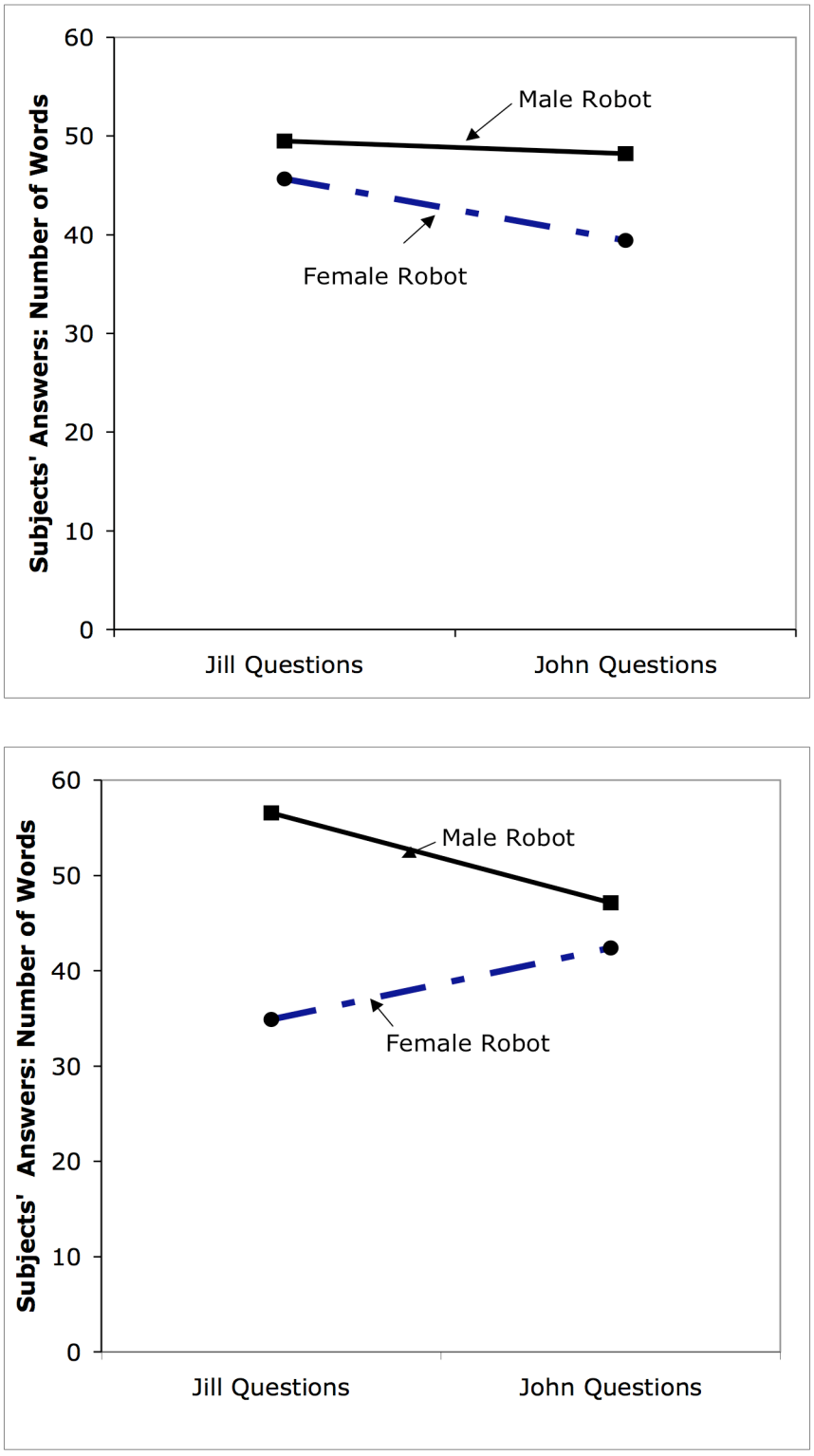

Figure 2. Number of words participants used in answering a male or female robot's questions about a woman (Jill) or a man (John) in a dating scenario (from Powers et al., 2005).

Figure 3. Our robot as a cooking assistant, programmed with knowledge of cooking tools (from Torrey et al., 2006).

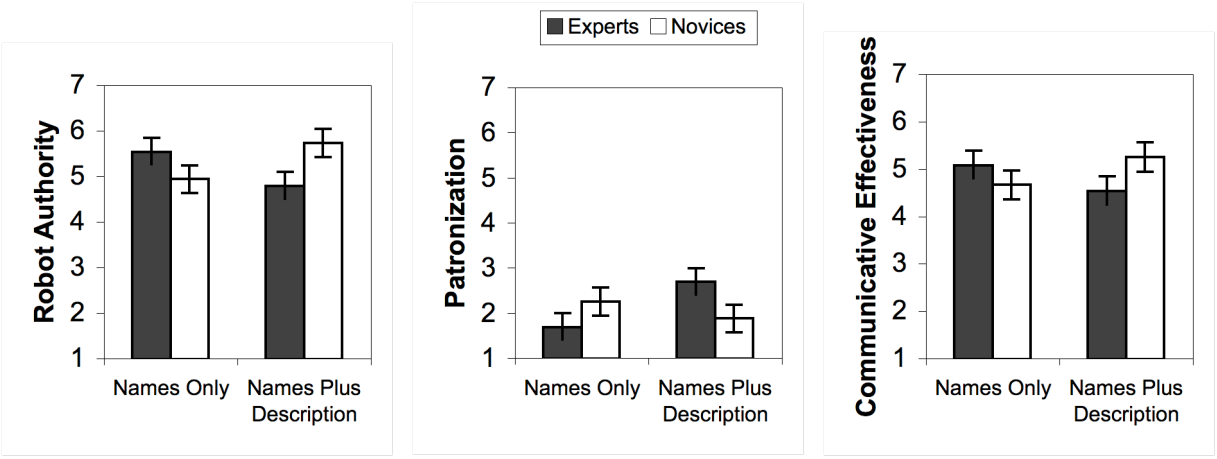

Figure 4. Experts and novices evaluate the robot more positively when the dialogue is adaptive to their information needs (from Torrey et al., 2006

Figure 5. On top of the monitor where the cooking tools are displayed, a small, eye contact sensor was mounted to detect whether the participant is looking in the direction of the monitor (from Torrey et al., 2007).