How Robots Can Help

Communication Strategies that Improve Social Outcomes

Cristen Torrey

May 2009

CMU-HCII-09-102

Human Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Thesis Committee:
Sara Kiesler (Co-chair), Carnegie Mellon University
Susan R. Fussell (Co-chair), Cornell University
Anind K. Dey, Carnegie Mellon University
Jeffrey T. Hancock, Cornell University

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

This work was supported by the National Science Foundation under grants DGE-0333420, IIS-0121426, and IIS-0624275. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect those of the funding agencies.

Keywords

Human-computer interaction, human-robot interaction, social robots, robot helpers, assistive robotics, robot assistants, human-robot communication, conversational interfaces, adaptive dialogue, help giving communication, help messages, help giving, social responses to technology, computers as social actors theory, perspective taking, politeness theory, help seeking, reactions to aid, informal learning, just-in-time learning, social and emotional outcomes, expectancy violation theory, linguistic mitigation, face threat, hedges, discourse markers, baking instruction.

Abstract

Offering help is a socially precarious venture. A robot that mimics human help-giving communication might end up supporting or might end up offending the help recipient. This thesis begins by observing the varied linguistic strategies human help givers use and their subsequent effects on help recipients. With this understanding, this thesis experimentally observes reactions to robot helpers in comparison to human helpers, looking closely at the influence of help messages on impressions. This experiment provides evidence that imperative statements from a robot are perceived to be controlling, in much the same way as humans using imperative statements. But when particular politeness strategies are used, robots are judged to be even less controlling than people. This thesis improves our understanding of human help-giving communication, offers guidance in the design of sensitive robot helpers, and argues for the continued investigation of advantageous differences between social responses to technology and social responses to people.

Acknowledgments

I have spent a great deal of time in the past couple of years thinking about help-giving behavior. Thankfully, during my years at Carnegie Mellon, I have been surrounded by excellent examples of smart and sensitive helpers. My advisors, Susan Fussell and Sara Kiesler, are two of the brightest and most knowledgeable people I have known, and I am deeply appreciative of the time and energy they have invested in my intellectual development. My thesis committee members, Anind Dey and Jeffrey Hancock, each generously contributed their unique perspective to this work, and this thesis is the better for it.

My partner, David Holstius, helped me move baking supplies, test cupcake recipes, and patiently listened to me talk about robots at the dinner table. He has been a crucial part of my success, along with my parents Chris and Connie Torrey, my sister, Clarice, and my dear friends Bethany Armstrong and Carl DiSalvo, all of whom graciously supported and encouraged me throughout my journey in graduate school.

My research at Carnegie Mellon has benefited from the help of capable research assistants and collaborators, including Aaron Powers, Matthew Marge, Hau-Yu Wong, Andy Echenique, Levi Boyles, and the members of the Project on People and Robots. I also had the privilege of working with several undergraduate drama majors to create the video clips used in Experiment 4. Nina Demeter, Sara Trapnell, Arya Arabshahi, Evan Tyler Barron, Noah Plomgren, and Daniel Scott

Weschler were wonderful, and I greatly appreciate their willingness to participate in my little research production.

Finally, I would like to thank my office mates, Peter Scupelli, Bilge Mutlu, and Matthew Easterday, for their emotional and intellectual support over the years, and Queenie Kravitz, our graduate student coordinator, for taking care of me and trying to make sure I took care of myself. Thanks also to. my fellow HCII PhD students, past and present, for their friendship and laughter, especially Aruna Balakrishnan, Moira Burke, Scott Davidoff, Darren Gergle, Amy Hurst, Matthew Lee, Ido Roll, Irina Shklovski, and Ruth Wylie,

For David,

an irrepressible helper

Table of Contents

List of Figures

List of Tables

Chapter 1: Robots as Help Givers

Robots have a powerful pull on the public's imagination, particularly when conceptualized as domestic servants, assistants, or in other helper roles. In an imagined technical future, robots are intelligent and patient, able to assist people in a range of tasks. This vision of robots as helpers is being actively pursued in various academic and commercial research projects. Robots are guides in train stations (Hayaski, et al., 2007), reference librarians (Behan & O'Keefe, 2008), grocery shopping assistants for the blind (Gharpure & Kulyukin, 2008), rehabilitation exercise coaches (Kang & Mataric, 2005), personal assistants in the home (Walters, Dautenhahn, Woods, & Koay, 2007), and weight loss coaches (Kidd & Breazeal, 2007). This vision of robots as helpers is beginning to be available commercially, although still limited in its realization (Figure 1).

Figure 1. Robot helpers in commercial development: (from left to right) Nuvo by ZMP, Wakamaru by Mitsubishi, Papero by NEC, and Enon by Fujitsu

A range of technical challenges in the development of robot helpers certainly remain. This thesis explores this potential future, not by investigating what robot helpers will be able to do, but by investigating how human help recipients will respond. Successful interaction with robot helpers means achieving positive outcomes for people, outcomes that go beyond metrics like task efficiency. While improvements to task success or task efficiency may be elements of a successful outcome, using task success as a proxy for human success is incomplete, at best. To understand human responses to robot helpers, this thesis focuses on the social and emotional experiences of help recipients. By understanding human responses to and experiences with robot helpers, this thesis provides general guidance for the future development of assistive robotic technologies.

Robot Helpers

Robot helpers are designed to intentionally evoke social responses in people, through some combination of human form, movement, and natural language. In a recent conceptual outline of human-robot interaction research, Dautenhahn argues that social skills are an essential part of how robot companions are defined (Dautenhahn, 2007). In order to be considered useful, a robot helper must provide assistance in socially acceptable ways. Developing a socially acceptable robot is a rather vague notion, but a common approach in the human-robot interaction literature is to mimic human behavior. It is frequently hypothesized that mimicking aspects of human interaction will make the robot seem familiar and acceptable, contributing to task success and improving technology adoption.

Many robot helpers evoke elements of the human form, though often in a highly stylized way. For example, the robot RI-MAN (Figure 2) uses an abstract humanoid form to communicate its capacity to lift and to carry people in what is intended to be a non-threatening manner. It was designed with certain characteristically human features while diverging in terms of color and texture; the robot's body is soft plush in order to carry the person comfortably in the robot's arms.

Figure 2. RI-MAN, a robot to lift and carry people. The Institute of Physical and Chemical Research at Nagoya, Japan.

Human-like movement and gesture has also been used to orient people to the robot's behavior in a way that is familiar to them. The way robots grasp and offer objects, as well as the way robots use gaze is thought to generate a set of expectations about how the robot will behave. The robot, Twendy One, is designed to assist people living in wheelchairs in a variety of daily tasks around the home. The robot extends its arm to offer support for getting in and out of the wheelchair, as a person would do, and grasps objects such as trays (Figure 3) with four of its fingers, as a person would do. There are other ways, perhaps technically easier ways, a robot might engage in these activities, but the metaphor of a human care giver is a compelling one for designing interaction behavior. Because the robot mimics human movement, the designers of Twendy-One believe that people interacting with the robot will be able to anticipate the robot's actions and infer the robot's intentions.

Figure 3. Twendy One, a robot for support of activities of daily living. Sugano Laboratory at Waseda University.

In addition to human form and human movement, roboticists use natural language, as appropriate to the task, as a way of informing and directing, as well as soliciting information from their listeners. Increasingly, the design of robot helpers' communication behavior rejects strategies traditionally associated with machines, such as a specialized interaction vocabulary (for example, using the word "affirmative" to confirm commands), in favor of more sophisticated, socially supportive strategies modeled after natural human speech. These robots want to support their listeners' goals and persuade listeners to accept their advice. For example, a robotic rehabilitation exercise coach (Figure 4) encourages compliance with exercise goals by using phrases such as, "Good to see you're using your arm. That's perfect." A robot that praises its listener is hypothesized to motivate the listener to continue the exercises. A weight loss coach robot, called Autom (Figure 4), invites people to record their daily exercise by saying, "It is also helpful for us if you tell me how much exercise you got today. Will you do that now?" This robot is asking a personal question about how much the listener exercised, so it phrases this request politely. With a range of verbal strategies, such as praise or reference to a shared goal, these robots seek to generate positive social rapport.

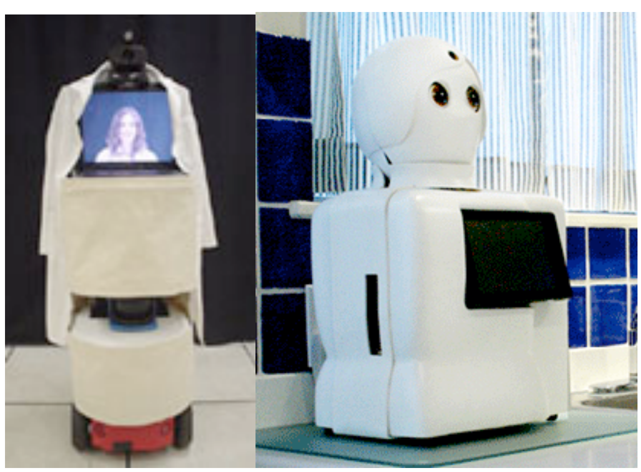

Figure 4. Clara (on left), an exercise coach. Robotics Research Lab at University of Southern California. Autom (on right), a weight loss coach. Media Lab at MIT.

Robot helpers may imitate different aspects of human behavior in different ways, but they share the following assumption: Modeling robot helpers on human form, movement, and language will facilitate interaction and increase compliance. But careful investigation is necessary before we can reliably use social responses to robots to generate desirable outcomes, particularly in the general task domain of help giving. Social psychological research on receiving aid from human helpers describes a range of influential factors and a mix of positive and negative emotional outcomes (Fisher, Nadler, & Whitcher-Alagna, 1982). Help recipients may perceive a help giver is compassionate and feel good about their success, or they may feel a help giver is being judgmental and feel their control over their activity has been threatened. Existing human-robot interaction research seeks the benefits of using human social cues, to increase familiarity, liking, trust, or persuasiveness, while ignoring the extensive range of other social responses that could be elicited. As just one example, it is possible that a robot functions as an audience for low-performing people, making them anxious about evaluation and causing them to perform even more poorly than they would if they were alone, a phenomenon that can occur in the presence of a human audience (Cottrell, Wack, Sekerak, & Rittle, 1968). Understanding social responses should include attending to the range of potential responses a robot helper could inadvertently elicit, including negative responses such as shame, anxiety, or coercion.

Reactions to Receiving Help

Investigating the full range of social responses to robotic helpers is particularly important because receiving help can have both positive and negative impact. Receiving help can certainly improve task success or efficiency; receiving help can even create or strengthen social bonds between the help giver and the help recipient. At the same time, help recipients may feel vulnerable—that their self-esteem and their control over their activity have been threatened. Accepting help implies that recipients are no longer completely responsible for the products of their activity. Research shows that help recipients may be concerned they will lose credit for a successful outcome, particularly damaging if they are highly invested in the task (DePaulo & Fisher, 1980). To protect their ownership, people may prefer to struggle with a problem rather than seek help (F. Lee, 1997).

Help recipients may also fear the perception that they are incompetent and dependent on others (F. Lee, 2002). Accepting help may require people to modify their self-perceptions, perhaps to wonder if they are as capable as they had thought. Other people's perceptions of the help recipient are also important. Help recipients can feel evaluated by their helpers; they may even resist being seen receiving help to avoid negative social judgments from bystanders.

Another source of vulnerability for help recipients is the potential for indebtedness (Greenberg & Westcott, 1983). When there is no opportunity to reciprocate, people are less likely to seek help (Nadler, Peri, & Chemerinski, 1985). Receiving help is more comfortable when the help recipient believes there will be an opportunity to help the other in turn (Flynn, 2006). These social costs—loss of task responsibility, negative estimations of competency, and indebtedness—could also play a role in interactions with robotic helpers, making helping interactions a highly sensitized area for studying the impact of modeling robot behavior on humans.

While the impact of receiving help can be mixed, it is certainly true that not all help is received poorly. Some help givers appear to be quite good at supporting a help recipient's social and emotional needs. So what mechanisms mediate positive and negative reactions to help? According to the "threat to self-esteem model," clusters of positive and negative reactions to help are driven by the extent to which the help supports or threatens the recipient's sense of self (Fisher, et al., 1982). A threat to self-esteem can exist simply by virtue of the help giver's characteristics. For example, it is more threatening for an adult to ask a child for help than another adult, so less help is sought from children (Drurian & DePaulo, 1977). People are also less likely to seek help from physically attractive people as compared to less attractive people (Nadler, 1980). Particular kinds of tasks can also be threatening for people because they are central to their identity (Nadler, 1990). So someone who sees himself as a dancer is more likely to interpret help with his dancing technique as threatening than someone who is out on the dance floor for fun. The way help recipients want to see themselves in relation to the task has an impact on their receptiveness to help.

Impact of the Help Message

There are characteristics of the task, the help giver, and the help recipient that can influence the extent to which help is perceived as threatening, but what about the help giver's behavior (or the robot help giver's behavior)? There are also aspects of the help giver's behavior that can impact the recipient's reaction positively or negatively. The help message itself can contain elements that threaten or support self-esteem. If the help giver emphasizes the help recipient's inferiority, the help can be seen as threatening. Or if the help giver feels it would be simpler to just complete the task on behalf of the recipient, the recipient may feel dependent and lack the confidence to make an independent attempt in the future. In contrast, help that supports the recipient's self-esteem might emphasize that the help recipient is encountering a common difficulty, one that many other people have. The work by Nadler, Fisher, DePaulo, and others suggests that the help message could influence the recipient's reaction by supporting or threatening self-esteem, but it does not test specific help messages empirically to determine their effect.

Another line of research, on the linguistic construction of requests, outlines further elements of a socially sensitive message. Politeness theory (Brown & Levinson, 1987) describes different ways speakers construct messages to support positive politeness, reinforcing personal connection, and negative politeness, avoiding threats to listeners' autonomy. For example, instead of asking someone to "take the kettle off the stove" an appeal to negative politeness would take another form, such as "I think the water might be boiling." Politeness theory proposes that people use these mitigating linguistic forms in order to avoid giving orders to other people, a very threatening action. Although politeness theory specifically describes the linguistic features around making a request, the linguistic features it outlines may be relevant to help giving as well. Giving help or advice is not, strictly speaking, a request, but the act makes a similar threat to the listener's autonomy, so one might expect successful help givers to use politeness cues to mitigate the force of their language.

Politeness theory offers greater detail about the specific language features involved in generating a supportive help-giving message. Brown and Levinson describe fifteen negative politeness strategies and fifteen negative politeness strategies (Brown & Levinson, 1987). Although these strategies are analyzed in isolation, they frequently co-occur, making it difficult to predict which combinations of politeness features would be appropriate for helping interactions. One difficulty lies in knowing how many polite features to use. It certainly seems possible that too many of these politeness cues would backfire, by saying something like, "excuse me, if you don't mind, I was wondering if there was just a chance that you might close the window." Research on politeness has not empirically evaluated whether this statement is polite or patronizing. Politeness theory offers a good framework for understanding a number of specific linguistic approaches that people use to manage potential threats to autonomy in conversation, but it is not a complete prescriptive understanding of how to generate positive reactions to receiving help.

Prior research on human help giving describes a range of possible reactions to help including positive effects like feeling connected and cared for and negative effects like threats to autonomy and competence. The literature suggests a number of factors that influence the help recipient's reaction to help. This work focuses specifically on the help message itself, the specific linguistic strategies that influence a help recipients' perception of the robot, attitude toward their task, and even beliefs about themselves and their own aptitude. By identifying successful strategies for a robot helpers' verbal behavior, this research stands to make a significant impact on help recipients' experience. There are relatively few technical obstacles to modifying the words a robot uses, so the robot's language use is an attractive opportunity to influence the development of robot helpers toward the needs of help recipients.

Thesis Overview

Because robot helpers interact with people, it is widely acknowledged that robot helpers must be socially intelligent in order to be successful. To that end, the design of robot helpers has attempted to mimic various aspects of human behavior in order to benefit from the social responses this behavior engenders. But offering help is a socially precarious venture, a situation where mimicking human help givers is not guaranteed to produce positive outcomes for help recipients. The goal of this thesis is to clarify the behavior and assumed role of robot helpers, as a class of assistive technology, by emphasizing the impact of conversational behavior on the help recipients' social and emotional responses.

This thesis begins by observing the varied linguistic strategies human help givers use and their subsequent effects on help recipients. With this understanding, this thesis then seeks to understand reactions to robot helpers in comparison with reactions to human helpers, looking closely at the construction of help messages as a source of influence on impressions. This investigation increases our understanding of human help-giving behavior and contributes to the research community's awareness of the social and emotional dimensions of interacting with assistive technology.

This thesis takes the following approach to investigating social and emotional responses to a robot's help message:

- In Chapter 2, I summarize existing empirical evidence showing that human responses to technology, from televisions to robots, mirror human responses to other humans. The most widely cited theory in this vein (Computers as Social Actors Theory, CASA) has been appropriated in the field of human-robot interaction to argue that modeling a robot's behavior after human behavior will generate positive outcomes. This chapter reviews the empirical evidence and reflects on the mixed results for this claim.

- In Chapter 3, I describe two experiments consistent with the dominant approach described in Chapter 2. In this series of experiments, I tested the social benefits and consequences of a robot with the ability to tailor its communication to its listeners. These experiments attempt to measure participants' social and emotional reactions but offer inconclusive results overall about the social benefits of perspective-taking strategies. An improved research paradigm, I argue, uses verbal manipulations observed in naturally occurring speech and uses a human control condition.

- In Chapter 4, I review direct observations of human help-giving communication, in the field and in the laboratory, undertaken to remedy concerns with the approach of the initial experiments. This work details several linguistic strategies used by help givers to avoid threatening their recipients' control over the activity. These observations also highlight several unexamined issues in the assumption that human behavior is the appropriate model in the development of robots.

- In Chapter 5, I describe the results of an experiment investigating reactions to help-giving communication using hedges and discourse markers. I based my linguistic manipulations on observed human speech, and I directly compared perceptions of human helpers with perceptions of robot helpers. This experiment demonstrates an alignment between perceptions of human and robot speakers on the negative effects of direct help messages; these messages are perceived to be more controlling. There is some indication that polite strategies such as discourse markers have a stronger positive impact for robot speakers than for human speakers.

- In Chapter 6, I conclude with a summary of findings related to three general areas of contribution. First, this thesis contributes to current understandings of human communication behavior. Second, this thesis explores reactions to the concept of a robot helper, emphasizing important but unexamined social and emotional responses. Finally, the approach of this thesis allows me to reflect on existing theory and research practice in the field of human-robot interaction and to further refine the dominate understanding of social intelligence in the field.

Summary of Research Contributions

First, this thesis contributes to current understandings of human communication behavior. It is well known that elements of the help giver's message are able to influence perceptions of the help giver and to influence the help recipient's attitude toward the interaction, but specific linguistic elements remain unexamined. This thesis empirically observes reactions to the use of hedges and discourse markers, as well as their combination. The use of hedges did improve perceptions of the speaker. While not typically recognized as a politeness strategy, the use of discourse markers improved perceptions in much the same way as hedges. Additionally, this investigation finds that there is a threshold for mitigating one's help message. The combination of hedges and discourse markers was not more successful than either one alone.

Second, this thesis draws attention to the importance of social and emotional responses to help-giving communication and offers guidance for the implementation of a robot helper's conversational behavior. This thesis contributes to our understanding of how robot helpers can generate positive social and emotional responses with careful attention to their help messages. Experiment 3 provides evidence that direct commands by a robot can be perceived as controlling, in much the same way as humans using direct commands. This thesis also suggests that robots may have a unique role to play as helpers. Robots using discourse markers were perceived as even less controlling than people. Expressions of politeness by robots can make a strong impression.

And, finally, this thesis examines two common assumptions in existing human-robot interaction research. First, this thesis questions the implicit assumption that social responses to robots are equivalent to social responses to humans. By observing a human control case, this thesis finds interesting opportunities where responses to humans and robots are not identical. Second, this thesis questions the implicit assumption that modeling a robot's behavior on human behavior will produce positive outcomes. Simply advocating the implementation of social behavior is not sufficient for the design of successful human-robot interaction. The case of help giving is a concrete example of the range of positive and negative outcomes that potentially follow from human behavior. Successful or not, human help-giving behavior is social. The questioning of these assumptions paves the way for a more sophisticated conceptualization of social intelligence, one that considers the complexity of human emotional responses and the range of opportunities in the design of a robot's behavior.

Chapter 2: Existing Evidence of Social Responses to Technology

Robot helpers mimic different aspects of human behavior, and research on reactions to human help givers reveals the potential for negative consequences, such as threats to independence or insults to the recipient's competence. But is it possible for a robot to generate these kind of negative responses? It is conceivable that help given by robots is simply not comparable to help given by human helpers. Receiving help from a robot might not preclude the recipient from taking full responsibility, therefore receiving help from a robot would not threaten the recipient's independence. Or people may not feel evaluated by a robot that offers help; the robot's negative judgment would have little impact. If robots could avoid all negative outcomes, they would be truly be ideal helpers, but there remains cause to be skeptical about the chances that robotic helpers (especially those that use language) will be perceived radically differently from human helpers.

Research in human-robot interaction (and human-computer interaction more broadly) consistently demonstrates a willingness on the part of humans to interact with machines much as though they were other people. Human social responses to computers have fascinated researchers for some time. The most influential body of work in this area is summarized in *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places* (Reeves & Nass, 1996). This book describes a number of seemingly unlikely ways people respond to computers as though they were other people. Following up on this "computers as social actors theory," researchers interested in interactions with conversational agents, pedagogical agents, and robots have investigated the ways social cues might be used by technology to provoke desirable human responses. These attempts have had mixed results. In this chapter, I review existing empirical work investigating social responses to technology (and to robots, specifically, whenever possible).

Robots as Social Actors

The earliest studies of social responses to computers use a straightforward paradigm—take an outcome from a study in human-human interaction and replace a human in the situation with a computer. Using this approach, researchers have found that people respond to computers as though they have gender (E. J. Lee, Nass, & Brave, 2000), an introverted or extroverted personality (Nass & Lee, 2001), a sense of humor (Morkes, Kernal, & Nass, 1998), and the ability to flatter (Fogg & Nass, 1997).

Studies of robots have demonstrated several similar effects. In one study, participants were asked to estimate the likelihood that a robot would recognize different landmarks (S. Lee, Kiesler, Lau, & Chiu, 2005). 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. This experiment suggests people estimate a robot's knowledge based on its "nationality" in much the same way they do with other people (Fussell & Krauss, 1992; Isaacs & Clark, 1987). In another study, participants were asked to inform a robot about the modern rules of dating (Powers, et al., 2005). When the robot had grey lips and a male voice, participants offered more detail in their responses to questions like who should do the planning and who should buy new clothes for a first date. Participants offered more information to a "male" robot presumably because they estimated the "male" robot required more information on this topic. This experiment suggests participants use human gender stereotypes to construct messages to a robot.

As a body of research, these experiments (as well as numerous others) provide support for the claim that certain social patterns observed among people are similarly observed between people and robots. The claim that people engage with robots in ways that are similar to other people has been used to articulate a particular view of social intelligence, one that sees human behavior as the definitive model of desirable social interaction. If humans are going to inevitably treat robots as though they were people, this argument offers, then we should build robots to correspond to this expectation. This approach acknowledges the call by many influential roboticists for socially intelligent robots, for example (Breazeal, 2003; Fong, Nourbakhsh, & Dautenhahn, 2003), but it does so with a particular view of social intelligence, one that sees mimicking human behavior as the appropriate and necessary direction for models of social intelligence.

The claim that robots are social actors, much as people are social actors, is being used by some robot designers and developers to improve the outcomes of human-robot interaction. These researchers seek to implement human conversational behavior on a robot because it will motivate people to comply with the robot's direction, to lose weight for example or do their rehabilitation exercises. This motivation, it should be noted, is substantially different from those researchers attempting to develop increasingly human-like robots as an exploration of what it means to be human (Kahn Jr., et al., 2007; MacDorman & Cowley, 2006). At its core, this is a philosophical query about understanding human behavior and understanding the boundaries between humans and machines. For example, Freier observes human responses to a robot that makes claims about its moral agency; the robot claims its rights have been violated when, instead of receiving its next turn during a game, it is put into a closet (Freier, 2008). Research like this is not attempting to engineer more socially appropriate robots; it is exploring the moral claims a robot might make and their subsequent impact on people. These attempts to imitate human behavior in human-robot interaction research operate under a substantially different imperative than the exploration described in this thesis. The primary concern of this thesis is to explore the appeal of human behavior as a gold standard for researchers interested in engineering successful robots.

Using Human-Likeness to Improve Social Outcomes

As previously discussed, a range of social responses to robots have been observed. But using these responses to achieve positive outcomes for people is more challenging than it may appear. Capturing the mechanisms of human social behavior for the purposes of directly influencing impressions and actions of people can be difficult. Positive responses to a robot modeled on human behavior has been observed for a range of physical behaviors including approaches (Satake, et al., 2009), body orientation to express attention(Yamaoka, Kanda, Ishiguro, & Hagita, 2009; K. Yamazaki, et al., 2007), gaze behavior (Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009; Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009; Staudte & Crocker, 2009), and head movements (A. Yamazaki, et al., 2008). These models of physical human behavior allow the robot to move in familiar ways that communicate the robot's intention, particularly to initiate or engage in conversation.

Attempts to model other kinds of human behavior, such as personality or affective gesture, have met with more limited success. Gockley and colleagues created a "moody" robot; they hypothesized that people would be more likely to initiate conversations with the happier version of the robot and would communicate longer with the happier robot. Contrary to their predictions, people appear to be more interested in talking with the sad robot (Gockley, Forlizzi, & Simmons, 2006). Manipulations of a robot's mood or personality are frequently simplistic in nature, but, even so, sometimes these simple representations of human affect or personality do achieve the desired outcomes. Other research projects have had greater success by matching the robot's personality characteristics to the personality of the listener. Tapus et. al. describe an experiment in which a robot designed to encourage post-stroke rehabilitation exercises adjusted its behavior to correspond with whether the patient tested as introverted or extraverted (Tapus,

Tapus, & Mataric, 2008). The introverted robot made nurturing comments such as, "I know it's hard but remember it's for your own good," while the extraverted robot made comments like, "You can do it! Concentrate on your exercise." These manipulations, when matched to the personality of the patient, were successful in increasing the amount of time that patients spent with the robot, doing their exercises.

Within the related domain of conversational agents, using models of human behavior to generate desirable outcomes has met with mixed results as well. In their research on conversational agents, Cassell and Bickmore suggest that trust in interactions with conversational agents can be developed via small talk, such as when a realtor says, "Boston is certainly more expensive than it used to be" (Cassell & Bickmore, 2003). Their model of social language predicts that small talk avoids face threats, strengthens reciprocal appreciation, builds common ground, and develops coordination between the human and the conversational agent. An evaluation of REA, an embodied conversational agent, compared a task-oriented dialogue to a dialogue that also contained small talk. Small talk increased extroverted participants' ratings of trust but did not effect introverted participants' ratings of trust (Cassell & Bickmore, 2003). Small talk, as a range of linguistic activities, did not have a consistent positive effect on all kinds of participants. While the use of small talk may have benefits, impressions of small talk are more complicated than their current implementation had anticipated.

A later study, lasting several weeks, tested relational dialogue behavior in the context of a conversational fitness advisor. In the relational condition, the agent used humor, empathy, social dialogue, politeness strategies and other verbal behavior to develop the relationship to the human participant. In contrast to a non-relational control, the relational agent was rated higher on measures of liking, trust, and respect (Bickmore & Picard, 2005). Because the agent interacted with participants for such a long time, it was important that the agent have a range of different conversational strategies to maintain engagement, but this design is not able to isolate the outcome of any particular kind of conversational feature.

Conversational agents have also become a popular approach in the design of instructional online environments. Empathetic messages delivered by a conversational agent have been shown to be more successful at alleviating learner frustration than apologetic messages (Baylor, Warren, Park, Shen, & Perez, 2005). But participants reported being even less frustrated when the agent did not acknowledge the participants' frustration at all and instead remained silent. An agent using a variety of positive and negative politeness cues, such as indirect requests, was implemented in another instructional environment, and students learned more when interacting with the polite agent (Wang et al., 2008). These politeness cues were particularly effective when students reported a preference for indirect feedback. Designing dialogue for virtual tutors may not be quite that simple; indirect feedback may not always be the appropriate conversational approach. Following a fine-grained study of human tutoring communication, Person et. al. suggest the need for a tutor to explicitly assess a student's answer may outweigh everyday politeness strategies (Person, Kreuz, Zwaan, & Graesser, 1995).

Summary of Related Work

Consistent evidence has shown that social responses to machines exist, though it should be noted, however, that this does not imply they are equivalent. Nonetheless, the evidence for social responses to technologies like robots has focused the attention of designers of robot behavior on how these social responses should be managed, to ensure that they are generating desirable outcomes. Though it is not used exclusively, a consistent approach in the field of human-robot interaction is to base a robot's behavior on human behavior. As the number of attempts in this vein grow, however, it becomes clear that identifying appropriate human behavior and implementing it in a robotic platform is easier said than done. It is difficult to interpret human behavior, particularly communication behavior, and codify it in a mechanistic way that produces the intended effects. Basing a robot's behavior on appropriate human behavior is a start, but it is not nearly specific enough for generating behaviors like help-giving behaviors.

Chapter 3: Experimental Investigation of Perspective-Taking Strategies

Consistent with the prior work described in Chapter 2, my initial approach to investigating social responses to robot helpers began with the implementation of conversational behavior hypothesized to be desirable, based on existing theory. In the series of experiments described here in Chapter 3, the verbal manipulations are based on the literature on perspective taking. Perspective-taking behavior among human interlocutors has been described by a significant literature, and it seems likely that interactions with robots would benefit by adopting this behavior. Robot helpers may interact with people of varying levels of expertise, particularly when providing instruction, direction, or other kinds of information. In order to be sensitive to the needs of their listeners, robots could adapt their messages. Listeners who would benefit from very detailed information would get that additional help, and listeners with prior knowledge would not have to listen to more information than is necessary. The use of perspective-taking strategies makes communication efficient and also allows help givers to avoid insulting their listeners with too much or too little information.

In the experiments that follow, I manipulated the robot's verbal perspective-taking behavior, observing reactions to the presence or absence of different perspective-taking strategies. In Experiments 1 and 2, the robot used a simple model of the user's expertise to adapt the help message. Subjective impressions of the robot were only affected when participants were motivated to finish the task quickly. In Experiment 3, the robot adapted the help given in response to the listeners' gaze and task activity, as indicators that further information was required. The presence of these strategies did not significantly influence the social outcomes that were measured in this experiment. In this chapter, I review the design of these experiments and reflect on the deficiencies of the general research approach that makes their interpretation inconclusive.

Perspective-Taking Theory

The theoretical motivation for these experiments comes from the literature on common ground and the grounding process that unfolds in conversation (Clark, 1996; Schober & Brennan, 2003). Taking the perspective of another person during conversation is a distinctly social ability that underlies shared empathy, meaning, and cooperation. 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 (Clark & Wilkes-Gibbes, 1986; Fussell & Krauss, 1992; Krauss, Vivekananthan, & Weinheimer, 1968). Speakers attend to their listeners' group memberships and likely areas of expertise as they construct their messages (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 (Gergle, Kraut, & Fussell, 2004b; Lockridge & Brennan, 2002; 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 (Clark & Wilkes-Gibbes, 1986; Krauss & Bricker, 1966; Krauss & Weinheimer, 1964, 1966). These adjustments produce more effective communication, whether in the context of a single message (Fussell & Krauss, 1989) or over the course of an ongoing conversation (Kraut, Lewis, & Swezey, 1982; Schober & Clark, 1989).

Clark and Wilkes-Gibbs (1986) have proposed the concept of "least collaborative effort" to explain why messages that are adaptive to a listener's level of expertise are more successful. With appropriate messages, listeners can simply say "ok" or otherwise indicate that they understand. In contrast, messages that are inappropriately adapted to listeners' expertise will require more overall effort by both parties. If the message is too detailed, if directions for an out-of-towner were given to a local resident for example, the speaker has put forth more effort than necessary. If the message is not detailed enough, if directions for a local resident were given to an out-of-towner, subsequent clarifying discussion will be necessary.

Adapting to one's audience not only improves communication efficiency, but may also help maintain positive affect between speakers and listeners. When too little information is provided, listeners may interpret the sparse information as a sign that the speaker has no concern for their needs. Similarly, when too much information is provided, listeners may feel insulted. In general, people are motivated to maintain each other's "face" or positive impression of themselves (Goffman, 1955). One way in which speakers do so is by providing listeners with the right amount of information for their needs. Communications that threaten face can lead to negative evaluations of a speaker (Holtgraves, 2002). Appropriate adaptation has further been shown to facilitate social coordination and have other broad-reaching benefits for interaction (Galinsky, Ku, & Wang, 2005; Giles, Coupland, & Coupland, 1991).

Building a Perspective-Taking Robot

For the purpose of understanding the impact of perspective-taking strategies on human-robot interaction, it is also necessary to know about failures in perspective-taking and how conversationalists cope with inaccurate or inadequate perspective-taking. To date, there has been little investigation of such issues in the literature on human communication, and it is not entirely clear how the principle of "least collaborative effort" should be instantiated in the design of human-robot communication. Although effort expended in a conversation between two people 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.

In order to investigate the costs and benefits of perspective-taking strategies in human-robot communication, it is necessary to instantiate these theories on a robotic platform. In the experiments that follow, perspective-taking was implemented in two different ways in the context of a referential communication task involving a set of cooking tools. The participants' goal was to select ten cooking tools needed to make a crème brûlée dessert. Participants selected the tool by clicking on the correct picture on a computer monitor (Figure 6 shows the images on display when the participant is asked to select a saucepan). Each of the ten tools was displayed separately alongside five incorrect tools. The robot conversationally led the participant through the task, requesting each of the tools in turn, and answering participants' questions. Participants could ask the robot as many questions as they wished.

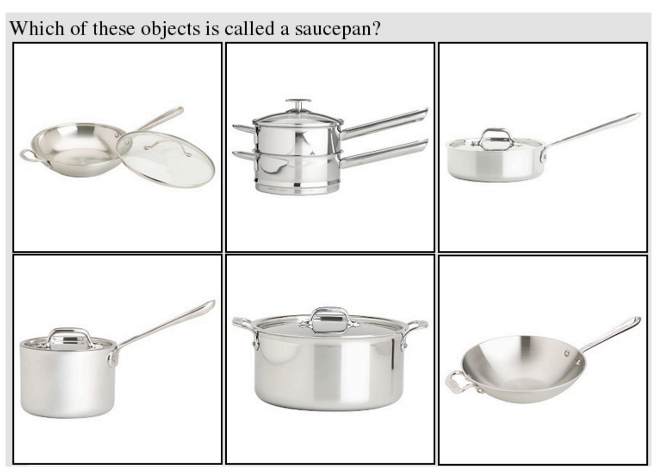

Figure 5. Screen display for cooking tool selection task.

Perspective-Taking Based on a Model of Expertise

The approach used in Experiments 1 and 2 is a user modeling approach. By user modeling, here, I 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. For the purposes of these experiments, I developed a pre-test to measure participants' cooking expertise. Sixteen pilot participants identified ten cooking tools from among a set of possible choices. These participants also answered eight questions in which they matched definitions of cooking methods with the names of these methods. Participants who could match the cooking methods with their definition were also able to correctly identify the cooking tools (r =.71, F [1, 15] = 15.4, p = .001). Potential participants for the subsequent experiments were asked only to match the cooking methods with their definition. Participants who scored 100% on this test were classified as experts. Those who scored less than 50% were classified as novices. 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.

Perspective-Taking Based on Listener Feedback

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. This is the basis of the strategy used in Experiment 3. 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-Gibbes, 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, 1978). 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, et al., 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, Kraut, & Fussell, 2004a). 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.

In Experiment 3, I implemented this approach in two different ways, by giving the robot simple awareness of the listeners' activity through gaze awareness and task activity awareness. The manipulation of gaze awareness I 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.

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, I describe how responses to these perspective-taking strategies were empirically observed in the laboratory.

Experiments 1 & 2: Adapted to a Model of Expertise

In Experiments 1 and 2, I examined the value of a perspective-taking strategy modeled on listener expertise by comparing a robot that uses language appropriately adapted to the knowledge of its listener with a robot that uses language that is not adapted to the knowledge of its listener. To identify individuals with different expertise, I developed a short test to categorize expert and novice cooks. We placed these novices and experts into two experimental conditions and observed their interactions as the robot asked participants to identify pictures of cooking tools used in making crème brûlée. In the Names Only condition, the robot simply named the cooking tools. This dialogue was appropriate for expert cooks. In the Names Plus Description condition, the robot described and explained the function of each tool. This dialogue was appropriate for novices. The design of Experiment 1 and Experiment 2 were identical, except that in Experiment 2 a financial incentive was offered to participants to encourage task speed and accuracy.

Method

The experiment was a 2X2 (expertise X dialogue) between-subjects design. We varied expertise, as previously described, by administering an online test prior to participation and selecting novices and experts for comparison. We created two dialogue conditions. In the first condition, the Names Only condition, the robot directed the participant to the tool by identifying the tool by name. This condition was hypothesized to be more suitable for experts. In the second condition, the Names Plus Description condition, the robot named the tool and further described it in several sentences. This condition was hypothesized to be more suitable for novices with little knowledge about the proper names of cooking tools (see Table 1 for example dialogue from each condition).

Table 1. Example directions for finding the paring knife.

Condition	**Robot Dialogue**
Names Only	Next you want a sharp paring knife. Find the paring knife.
Names Plus Description	Next you want a sharp paring knife. Find the paring knife. It's usually the smallest knife in a set. It has a short, pointed blade that is good for peeling fruit. The blade is smooth, not jagged.

Participants

Forty-nine students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University for Experiment 1. They were each paid $10 for their participation in this experiment. For Experiment 2, an additional forty-eight students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University. They were each paid $8 plus possible bonuses for speed and accuracy, up to $15 for participation in the experiment.

Robot

The robot used for this experiment was originally designed to interact with people in a nursing home (Pineau, Montemerlo, Pollack, Roy, & Thrun, 2002). In this experiment, the robot was stationary and was dressed to appear like a cooking expert. The robot wore a white chef's hat and apron and spoke with a male voice. The robot opened its eyes at the start of the experiment and closed them at the end. While speaking, the robot's lips moved in synchrony with its words. The robot's face is articulated and is capable of a range of expressions, but the full range of expression was not utilized in this experiment.

Procedure

When participants arrived at the experimental lab, the experimenter told the participant that the robot had been given "specific expertise" in cooking, and that "the robot will be talking to you about the tools needed to make a crème brûlée dessert." The robot spoke aloud and also displayed its messages on a display on the robot's chest. The robot used Cepstral's Theta for speech synthesis, and its lips moved as it spoke. The text also showed on the screen, as in Instant Messenger interfaces. Participants interacted with the robot by typing into the same Instant Messaging interface.

In the course of the dialogue, the robot prompted the participant to find cooking tools, e.g., "Find the picture of the saucepan." The tools were shown on a nearby computer (see Figure 7). If the participants knew which tool was correct, they clicked the correct image and told the robot that they found the right tool. If the participant did not recognize the name of the cooking tool, they could ask the robot questions about the tool, using the IM interface (some participants spoke out loud as well). Most of participants' questions were about tool properties like shape ("does it have a round bottom"), color ("what color is it"), and usage ("what is it for"). The robot was programmed to respond to most of these inquiries.

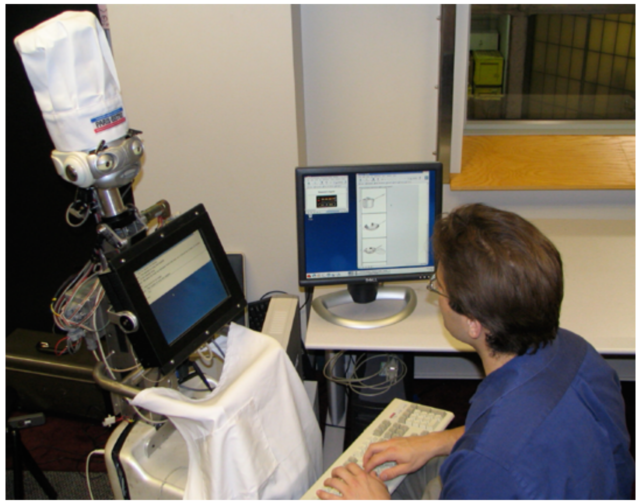

Figure 6. The set-up for Experiments 1, 2 & 3.

To create time pressure (during Experiment 2 only), participants were informed in the written instructions that if they finished the task quickly, they would receive an additional $1 for every minute under the average participant time. The experimenter answered any questions about the experiment and started a timer when the participant typed "hello" to the robot to begin the task. There was also an incentive for accuracy. If participants correctly identified all ten items, they would receive an additional $3 in payment. The experimenter displayed a running timer on the monitor where the participants were selecting the cooking tools. When they began conversing with the robot, the experimenter started the timer, and it was visible the entire time they worked at the computer.

In both Experiment 1 and 2, all participants' responses to the robot were logged. After conversing with the robot, the participant completed a survey about their perceptions of the robot and their conversational interaction.

Measures

Interactions with the robot were measured on two dimensions, information exchange and social responses. In the course of telling the participant about making crème brûlée, the robot asked the participant to identify ten cooking tools. Information exchange was measured using the number of questions participants asked about the tools. A greater frequency indicates the participant did not know which tool was correct and needed more information. The number of tools correct was used as a measure of the accuracy of their performance. I did not expect large differences in performance, because the participants could keep asking the robot questions until they thought they understood which cooking tool was correct. I also measured the time each participant spent on the task and the number of misunderstandings with the robot.

Participants' social reaction to the robot was measured through self-report items on a questionnaire. Participants completed this questionnaire following their interaction with the robot. The questionnaire covered three general areas of interest: participants' impressions of the robot's personality and intellectual characteristics (authority, sociability, intelligence), participants' evaluation of the quality of the communication (effectiveness, responsiveness, control), and participants' evaluations of the task (enjoyability, ease).

Table 2. Scale reliability for measures of social relations.

Scale (Cronbach's α)	**Sample Item**
Robot Authority (0.72)	Expert/Inexpert
Robot Responsiveness (0.76)	My partner can adapt to changing situations.
Conversational Control (0.86)	My partner was more active in the conversation than I was.
Conversational Effectiveness (0.90)	I found the conversation to be very useful and helpful.
Task Difficulty (0.77)	This task was difficult.
Task Enjoyability (0.75)	I enjoyed participating in this task.
Robot Patronizing (0.90)	My partner's explanations can be condescending.
Content Appropriateness (0.72)	I got just the right amount of information from my partner.

To assess users' perception of the robot's authority (McCrosky, 1966) and sociability and intelligence (Warner & Sugarman, 1986), I used existing scales from the social psychology literature; I used these scales in their entirety. I also selected items from a published (but lengthy) communicative effectiveness scale (Canary & Spitzberg, 1987) and from a communicative competence scale (Weimann, 1977). I also created scales to measure task enjoyability and task difficulty, to measure whether the content was appropriate, and to measure whether the robot was perceived to be patronizing. I conducted factor analysis on the scales after collecting data in Experiment 1 to confirm that the scales were appropriate for evaluating a robot. The reliabilities for the scales are shown in Table 2.

Results

The robot asked participants to identify 10 cooking tools. The number of questions participants ask the robot is the measure of information exchange. The number of questions reflects the amount of uncertainty the participant has about which cooking tool is correct and is related to the amount of effort exerted by the participant in communicating with the robot. There was a significant interaction between expertise and the dialogue condition in both Experiment 1 (F [3, 48] = 9.99, p <.01) and Experiment 2 (F [3, 47] = 10.9, p <.01). Figure 8 charts the data from Experiment 1; the data from Experiment 2 show an identical pattern. These interactions show that novices were negatively impacted by the absence of additional detail in the Names Only condition. Without an appropriate amount of detail, novices had to work harder to communicate with the robot and get the information they required.

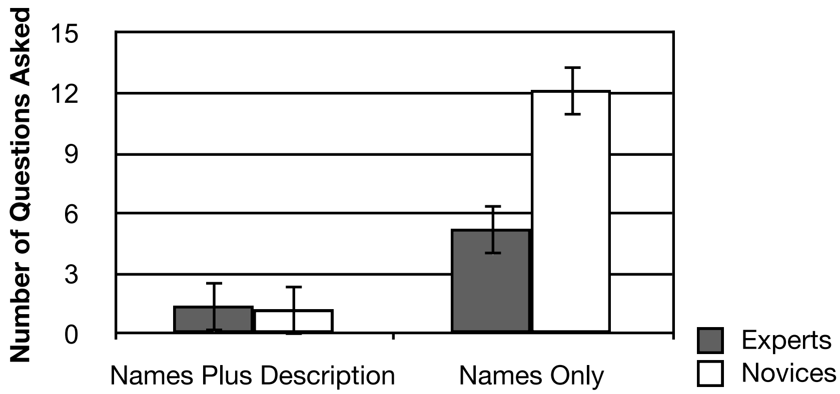

Figure 7. Novice users' performance was affected disproportionately by the robot's lack of description in Experiments 1 & 2.

Outcomes related to social relations were measured via the questionnaire data, but I observed significant interactions on these social measures in Experiment 2 only. In Experiment 2, we offered monetary incentives for speed and accuracy. It was only under these conditions that any positive and negative social impacts of perspective-taking behavior were observed. Three measures of social relationship produced a significant interaction (see Figure 9). First, the robot's authority was perceived differently depending on level of expertise and dialogue condition (F [3, 47] = 6.3, $p < .05$). Participants who conversed with the robot whose dialogue matched their level of expertise found the robot to be more authoritative than participants who conversed with a robot whose dialogue did not match their expertise. Thus, experts who interacted in the Names Only condition, and novices who interacted in the Names Plus Description condition thought the robot was more authoritative. Also, participants conversing with a robot whose dialogue matched their expertise thought the robot was less patronizing than a robot with mismatched dialogue (F [3, 47] = 4.5, $p < .05$). Finally, the questionnaire measure of communicative effectiveness, which included items like "Our conversation was successful," also showed a significant interaction (F [3, 47] = 10.97, $p < .01$).

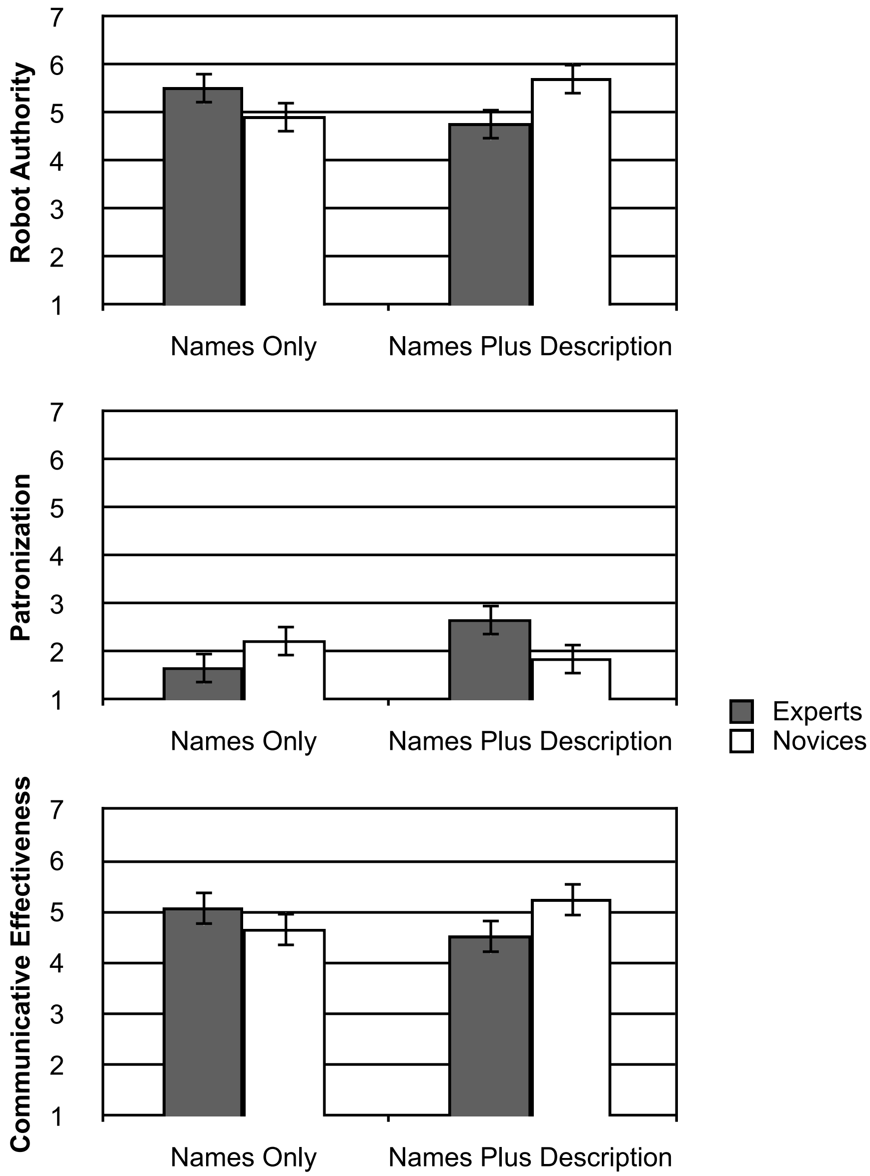

Figure 8. Under time pressure in Experiment 2, experts and novices evaluate the robot more positively when the dialogue is adapted to their information needs.

Discussion

In Experiments 1 and 2, I tested the impact of perspective-taking strategies in a very straightforward way. With a 2X2 design, I matched expert and novice cooks with conversational dialogue that was either a match for their expertise or a mismatch for their expertise. The impact of perspective-taking strategies was apparent with regard to the measures of information exchange. A lack of tailored information disproportionately impacted novice performance, but the social impact of this lack of information is far less clear. While it may seem obvious that human speakers lacking perspective-taking abilities would be perceived negatively, the robot lacking perspective-taking abilities was only perceived negatively when the experimental protocol emphasized speed and offered a monetary incentive for performance.

The implementation of perspective-taking based on a user model, created for Experiments 1 and 2, is a crude example of perspective-taking behavior. It places listeners into one of two categories, either people who need a lot of information about everything and people who do not need any additional information. Of course, one would hope that a human helper might differentiate between different listeners' expertise in a much more responsive, sophisticated way. In Experiment 3, I implemented perspective-taking behaviors that respond directly to the actions (and inaction) of individual participants.

Experiment 3. Adapting to Task Activity

In Experiment 3, I used the same general protocol as the one used in Experiments 1 and 2. The cooking tool selection task and the robot are exactly as previously described. The perspective-taking behavior implemented for Experiment 3 was designed to be responsive to listener feedback, specifically listener gaze activity and listener task activity. The robot makes an initial attempt by using the name of the tool and if the listener's gaze or task activity indicates this referent is insufficient for their level of expertise, then the robot elaborates on the name of the tool. In this way, the robot can provide information when the listener has offered evidence that further information is necessary. This approach has multiple benefits. First, it is easily adapted to listeners at any level of expertise and does not require any pre-testing from people interacting with the robot or the population at large. Second, the robot's awareness of and responsiveness to the listener's activity should indicate that the robot is considerate and willing to accommodate the listener's needs, potentially improving social responses to the robot helper.

Method

This exploration of the effects of a robot's responsiveness to human gaze and task activity included four trial conditions (see Table 3). In the baseline condition, the robot responded only to questions articulated by the participant. The next two conditions added increasing levels of awareness to the robot's capabilities with the addition of responsiveness to gaze and responsiveness to a delay in task progress. In the fourth and final condition, we isolated the effect of the delay by adding a condition which offered elaborations immediately, as each tool was introduced.

Table 3. The robot offers elaboration differently in each of the four trial conditions.

Condition	After Questions	After Gaze	After Delay	Immediately
Questions Only	✓			
Gaze Added	✓	✓		
Delay Added	✓	✓	✓	
Immediate Added	✓	✓		✓

Questions Only Condition

In the Questions Only Condition, the robot conversed with the participant, only offering additional information if the participant verbally requested it. For example, the robot introduced the paring knife by saying "Next you want a sharp paring knife. Find the paring knife." Participants who did not know which selection to make asked specific questions or told the robot they needed more information about the tool.

Gaze Added Condition

In the Gaze Added Condition participants were also able to ask questions of the robot. In addition, they were given further information about the tool they were attempting to select if they turned away from the computer display of the tools. Based on previous observations of participants, we assumed that people turned away from the task to look at the robot and ask a question. Thus, the robot elaborated on the tool when the robot sensed the participant had turned away from the task display. For example, if the participant was looking for the paring knife and turned toward the robot, the robot would respond, "The blade is smooth, not jagged."

Delay Added Condition

In the Delay Added Condition, participants received additional elaboration when they had asked a question or when they looked back at the robot, just as in the Gaze Added Condition. In addition to these opportunities, the robot in the Delay Added Condition provided hints when four seconds elapsed without a selection having taken place. Four seconds was the average amount of time it took a participant to select a tool in our previous experiment using the identical task. We thus assumed that when a participant hesitated for longer than four seconds, they were uncertain and could use further elaboration.

Immediately Added Condition

In the Immediate Added Condition, the robot offered additional elaboration when the participant asked a question or when the participant looked back at the robot, just as in the Gaze Added Condition. In the Immediate Added Condition, however, every time the robot introduced a new tool, the robot immediately added an additional elaboration about that tool. The robot did not wait for a delay, as in the Delay Added Condition; instead, the robot included a hint in its initial turn. In this condition, for instance, the robot introduced the paring knife by saying, "Next you want a sharp paring knife. Find the paring knife. The blade is smooth, not jagged."

Participants

Sixty-six students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University. They were each paid $10 for their participation.

Procedure

When participants arrived at the experimental lab, the experimenter adjusted the gaze sensing camera to the height of each participant. Participants were informed that their gaze was being tracked. The instructions given to the participant and the subsequent dialogue about the cooking tools with the robot proceeded as previously described in the procedure sections for Experiments 1 and 2.

Gaze Sensing Implementation

We implemented gaze sensing by mounting an eyeBox (xuuk.com) eye contact sensing camera (Smith, Vertegaal, & Sohn, 2005) to the top of the monitor where participants made their selection. The experimenter adjusted the monitor so that each participant's face was in the field of view of the eyeBox. Though it was only utilized in three of the four conditions, the eyeBox was introduced to and adjusted for participants in all conditions. The eyeBox uses infrared light to illuminate any pupils in the frame and outputs the number of eyes it finds in the frame at a rate of ten times per second. The output also includes whether the eyes were detected looking directly toward the camera and the location of the eyes in the frame. If two eyes were detected looking directly toward the camera, the robot registered eye contact.

Measures

In Experiment 3, I collected the following outcome measures: the participants' task performance, their verbal interaction with the robot, and their subjective evaluations of the robot, the conversation, and the task. As an experimental control, I also gave participants the pre-test for expertise. These measures were identical to those collected in Experiments 1 and 2.

Results

In this section, I explore the effects of both expertise and the condition manipulations on performance, communication, and subjective evaluation measures. Our model includes the elaboration condition, participant expertise, and the interaction between condition and expertise. In order to further determine which of the four conditions were significantly different from one another I compared all possible paired conditions using Student's t test.

Performance Measures
I first considered the effect of expertise and condition on the number of tools participants chose incorrectly before finding the correct tool (see Figure 11). There were two significant main effects and no interaction. Experts make significantly fewer mistakes than novices, $F(1,65) = 11.5$, $p = .001$ (Experts $M = 2.7$, $SD = 2.1$, Novices $M = 5.1$, $SD = 3.7$). The main effect of condition on selection mistakes is also significant, $F(3,63) = 2.98$, $p < .05$ (Question $M = 4.7$, $SD = 3.5$; Gaze $M = 5.3$, $SD = 3.5$; Delay $M = 3.5$, $SD = 3$; Immediate $M = 2.4$, $SD = 2.3$). Post hoc comparisons of all four conditions show that participants in the Immediate Added Condition do make fewer mistakes than participants in either the Questions Only or Gaze Added Conditions.

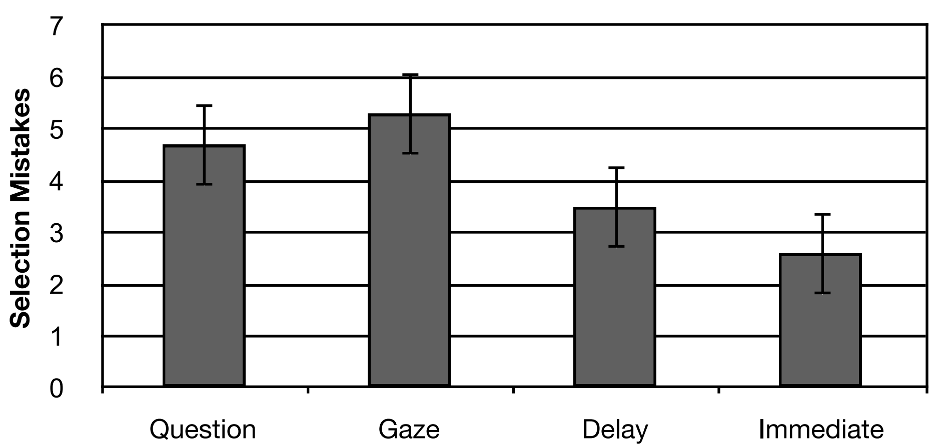

Figure 9. Experiment 3, number of selection mistakes by condition.

Next I considered the effects of expertise and condition on the amount of time participants required to complete the task. There was a main effect for expertise. Experts take less time to complete the task than novices, $F(1, 65) = 15.7$, $p < .001$. An ANOVA for the effect of experimental condition on time was not significant. Participants in each condition spent roughly the same amount of time conversing with the robot. There was no significant interaction.

As might be expected of those more familiar with the cooking tools, experts took less time on task and made fewer mistakes than novices. There was no interaction across conditions. There were few performance differences across participants in each of the four conditions, except for the finding that participants given immediate elaborations made fewer mistakes than those who had to request elaborations, either with questions or with their gaze.

Communication Measures

Next I considered the number of questions participants asked the robot. In addition to receiving elaborations as the result of a question, participants received elaborations immediately and/or as the result of gaze or task delay (depending on their condition). Figure 12 provides an overview of how many of each type of elaboration was received in each condition. Participants in the Delay Added and Immediate Added Conditions tended to receive a greater number of elaborations overall. The robot's awareness of gaze contributed only a small number of elaborations in each condition. As shown in Figure 6, the average number of elaborations participants received as the result of task delay was more (*M* = 7.1, *SD* = .6) than the number of elaborations participants received as the result of responsiveness to gaze (gaze-prompted elaborations in the Gaze Condition *M* = 1.75, *SD* = .4; Delay Condition *M* = .9, *SD* = .4, Immediate Condition *M* = 1.6, *SD* = .4).

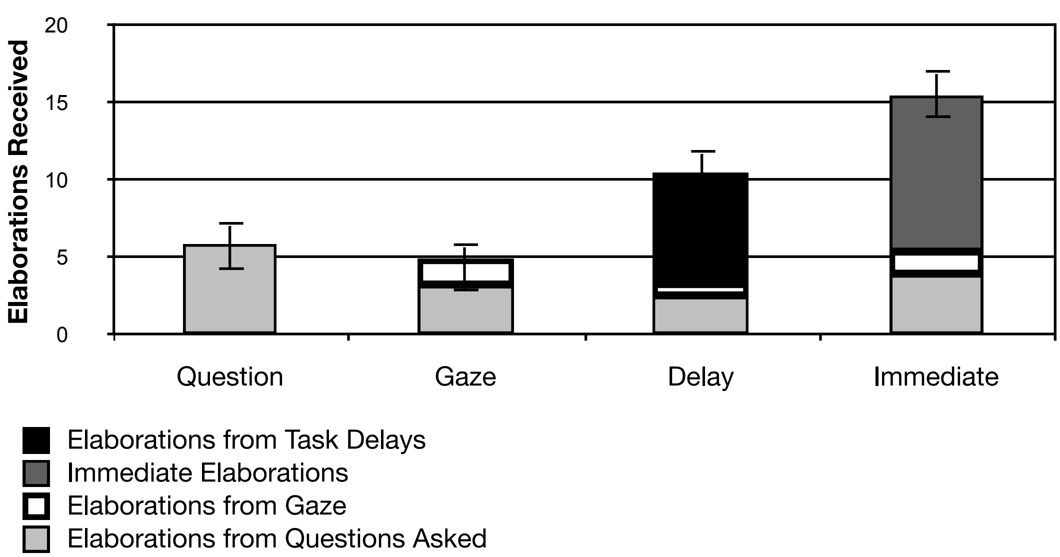

Figure 10. Experiment 3, the average number of elaborations received in each condition.

Our analysis of the number of questions participants ask in each condition again revealed a main effect for expertise; experts ask fewer questions of the robot than novices, $F(1,65) = 7.5$, $p < .01$ (Figure 13). There was no main effect for condition, and there was no significant interaction between condition and expertise. Student's t tests revealed that two conditions differed significantly from one another. The Questions Only Condition asked significantly more questions than the Delay Added Condition.

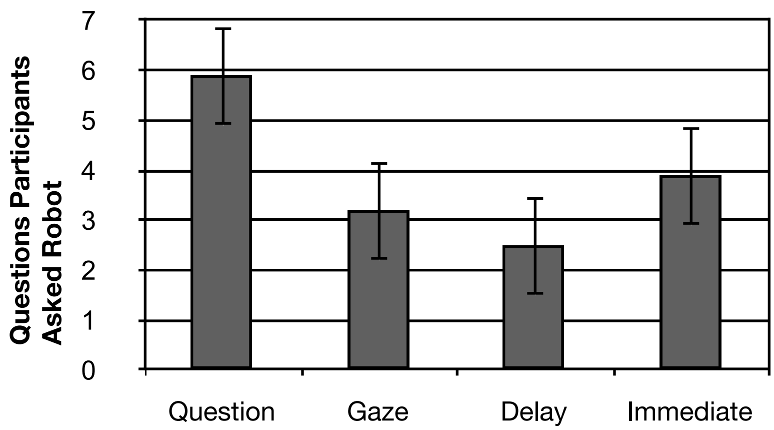

Figure 11. Experiment 3, number of questions participants asked by condition.

There were no interaction effects between expertise and condition on the communication measures, but there was a significant main effect for expertise. Experts ask fewer questions than novices. The only significant difference between conditions on the communication measures is a significant decrease in the number of questions asked when the robot is responsive to task delays compared to when the robot responds only to questions.

Subjective Evaluation Measures

There were no significant differences in the subjective evaluation measures by condition.

Discussion

Experiment 3 was designed to test two simple perspective-taking strategies, based on gaze activity and task delay. These strategies did increase the robot's responsiveness to participants' individual difficulties. Overall, the robot's awareness of task delay was more successful at getting participants more information when they appeared to require it. Comparatively few extra elaborations were the result of the robot's attention to gaze activity. But none of these strategies resulted in any significant changes to participants' perceptions of the robot. A robot offering additional information, if it became necessary, was not rated as any more likable, intelligent, or responsive by the participants in this experiment.

Summary of Experiments 1, 2 & 3

In Experiments 1, 2 and 3, I saw very little evidence that a robot using perspective-taking strategies will be perceived more positively along social dimensions. In Experiment 2, when participants were given a financial incentive to finish quickly, there was some evidence of social impact, but this laboratory context has the atmosphere of a contest and is not typical of one's everyday interactions. On the basis of these experiments, then, is it safe to conclude that perspective taking is not essential for robot communicators? Will human listeners easily overlook this social slight from robot speakers?

Unfortunately, the design of these experiments affords alternative interpretations. First, existing theory on perspective taking describes general conversational behavior and is not specified at a level of detail that can be implemented directly on a robot. Previous literature documents the evidence for perspective taking but does not articulate the relationship between the depth of elaboration required and the listener's level of expertise. In these experiments, I implemented several perspective-taking strategies that are compatible with existing theory, but it is possible that the way in which I specifically authored the robot's conversation is inaccurate or inadequate. This possibility implies that perspective-taking behavior may indeed have consequences for robots, but that the way it was implemented in these experiments obscures their impact. A true test of perspective taking would require a precise understanding of how to program this behavior.

In order to explore whether perspective taking is essential for robot communicators, one would also need to include a human control condition in the study design. The second alternative interpretation is that the participants in these studies would have responded to human speakers in much the same way. Based on prior work, we can anticipate negative reactions to human speakers that disregard the needs of their listeners, but without a human confederate as a control, it is impossible to interpret the current experiments.

The general paradigm for investigating social responses to robots isolates a pattern from existing social science research and tests it in the context of human-robot interaction. If this pattern holds, then the experiment is considered a success, and the dominant belief about robots (that responses to robots mimic responses to other humans) is further ratified. However, if the hypothesized pattern does not hold (as in Experiments 1, 2 and 3 just described) then the general paradigm makes it difficult to interpret these results. It is unclear whether the behavior was implemented appropriately, and it is impossible to know precisely how the measures in use would respond to a human speaker (or human confederate) without utilizing a human control condition in the experimental design. The remainder of this thesis addresses these central limitations of the general approach. In Chapter 4, I describe a series of qualitative observations I conducted in order to design behavior grounded in observable human communication behavior. In Chapter 5, I review the experiment I designed, including a human control condition, to compare perceptions of human and robot helpers directly, in a more conclusive way.

Chapter 4: Qualitative Observations of Help-Giving Communication

Many researchers have observed that speakers engage in indirect speech, particularly when navigating what Goffman calls a "face-threatening act" (Goffman, 1955). Offering someone help can be a face-threatening act, one that helpers navigate carefully to avoid insulting the recipient. Goffman refers to the delicacy observed in communication around face-threatening acts as "the language of hints" (Goffman, 1967). Fraser (1980), Caffi (1999), and others describe this communicative work as mitigation, a strategy for reducing the force of a message. Politeness theory is also extensively engaged in the description of indirect speech, as has already been noted (Brown & Levinson, 1987). Previous research has drawn our attention to a variety of ways that people protect the social and emotional needs of their listeners, in difficult social situations, by speaking indirectly. Although we can expect these issues to arise in the course of help-giving communication, existing theory does not prescribe communication at the level of specificity that could reasonably generate experimental manipulations. It is difficult to design a robot's communication behavior from theory alone. Numerous specific decisions about a robot's behavior could be responsible for experimental results, so it is unwise to make these decisions arbitrarily.

In order to provide a sound basis for the design of our robot helpers' communication behavior, I conducted a series of observations of human help-giving communication. In the course of these observations, my focus became tuned to how help givers mitigate their language use for the social and emotional benefits of their listeners, as I observed these needs to be a source of much of the linguistic variation between help givers. In the remainder of this chapter, I describe the protocol for my observations of human communication behavior. In discussing my initial findings from the first observation stage, I reflect on several unexamined challenges in basing a robot's behavior on observed human behavior. After observing a range of human communication behavior in both natural and exploratory settings, I detail linguistic features used by different help givers to mitigate the social and emotional impact of their help messages.

A priority for this work is to focus on a task with an agenda that is driven by the help recipient, not the help giver. When a help giver initiates the task and the sub-steps of the task, the help-giving role is closer to the role of an instructor. The vision of robotic helpers is for long-term, sustained interaction, quite different from a traditional, discreet instructional lesson. The autonomy of the help recipient, as well as the help recipient's ability to subjectively define success, is of primary importance. Because of these priorities, I wanted to use a task that would be authentically motivating and accessible, even for novice participants.

This research uses cupcake baking as an activity around which to study help giving and help seeking. Assisting people in baking activities is not the end goal of this research, but the baking task has a number of unique qualities. First, successful cupcakes are a subjective outcome. Some people may want their cupcakes to look perfect, but frequently people just want them to taste good, from their subjective viewpoint. A batch of cupcakes would be difficult to use as a performance measure. The baking task represents a class of activities that do not have an easily identified metric for success; instead, there is a range of acceptable outcomes. Second, while each cupcake recipe has specific directions, there are many different ways to complete the sub-steps. There are, of course, errors; adding salt instead of sugar, for example, will result in inedible cupcakes. Some of the steps in the recipe are flexible; others are less flexible. This makes giving and receiving help around the baking activity particularly interesting because the help offered might be useful without being mandatory. Third, eating cupcakes is enjoyable so participants are motivated to make good cupcakes. Because this research is primarily interested in observing baking as a subjective experience, it is very important that participants remain engaged in the task and in the interactions surrounding the task.

Baking in a Dormitory Kitchen

In order to expose myself to helping interactions firsthand, I organized three "cupcake nights" in an undergraduate dormitory on the Carnegie Mellon campus. The goal of these sessions was simply to observe communication behavior as people tried an activity that was new to them.

Procedure

After obtaining IRB approval, I put up fliers in the undergraduate dormitory advertising a "cupcake night" in the dormitory's downstairs kitchen. This dormitory commonly had community events in their kitchen, like making ice cream sundaes. At the first two baking events, I brought along baking experts. The baking experts were told to answer questions and to give the undergraduates help when necessary. At the final baking event, I brought several baking handbooks but did not bring a baking expert. During this session without a baking expert, I wanted to see if the undergraduates would assist one another more often. Interaction between undergraduates was observed. There was little, if any, consultation of the baking handbooks that were available.

All interactions in the kitchen were recorded with audio and video. As participants arrived, I briefed them about the recording equipment and asked them to review the consent form. During the cupcake events, I participated by baking a cupcake recipe of my own. At the first event, I nearly ruined the batch of cupcakes I made. This was not my intention, but it was advantageous because thereafter the participants did not appear to view me as a baking expert. Participants occasionally asked me where they could find ingredients or tools, but I was not a primary source of cupcake expertise in the dormitory kitchen.

Participants were offered a selection of eight recipes from which to choose. Participants worked on a recipe in pairs more often than they worked as individuals. Each event lasted about three hours. Participants came and went as they wanted, but at each event the group of participants waited together for everyone's cupcakes to come out. The bakers took their cupcakes back to their rooms, but it was very common for participants to share their cupcakes with the entire group in the kitchen.

Following the event, I reviewed the video and transcribed any communication where help was being asked for or offered.

Findings

When experts close to the participants' age offered help and when participants themselves offered help to one another, their messages seem to carefully avoid telling the help recipient directly what to do. In the following example, April refers to the information she will give as a "secret." By introducing the help in this way, she avoids passing judgment on the help recipient's lack of knowledge.

```
[Working over the mixing bowl, ADAM shakes baking powder
out of the can into the measuring spoon.]

BAKER - APRIL:    oh wait a minute

[APRIL takes the baking powder from ADAM's hands.]

BAKER - APRIL:    can I just tell you a secret? see this
                  thing? you go like this

[APRIL levels off the baking powder in the measuring spoon
with the metal lip of the baking powder container.]

BAKER - APRIL:    see? measures it off the top. you see
                  what I mean?
```

Help givers sometimes appeared disfluent, seeming to struggle with how to construct their messages. This is particularly interesting when the help being given is relatively straightforward and well within the help giver's range of expertise.

Instances like these suggest the help giver is managing the social constraints of the situation and is unwilling to be directive.

```
BAKER - ALEX:  is this firm or

EXPERT - ZACK: well if you over beat it it'll uh actually,
               that won't be good either. so pull it out
               for a second, if that's alright

BAKER - ALEX:  it says firm and glossy

EXPERT - ZACK: yeah, you're , that's , that's pretty good
```

There was one help giver in this sample that did not use a similarly indirect style. This expert was older than the participants and approached the situation as more of an authority figure. Her communication tended to be more direct; there was little opportunity for the help recipient to propose alternative actions.

```
EXPERT - DEE:     test one. watch your hands you don't burn

EXPERT - DEE:     no. not yet ... they smell good

BAKER - JAN:      but it's getting hard. I feel like I
                  always over bake stuff and I don't want
                  to-

EXPERT - DEE:     -that's why we're watching them
```

This direct style could certainly be more efficient; this expert approached her role as help giver by circling around the room giving quick tips and direct instruction. There were instances where participants resisted this intrusion by ignoring the instruction, by physically avoiding the expert, and by verbally rejecting the information.

This exploratory observation highlighted the ways that language use defines and reinforces the relationship between the help recipient and the help giver. The way that helpers offer help messages is influenced by their age, expertise, and other attributes but the messages instantiate and reinforce the relationship between the help giver and the help recipient—instructor and student baker, baker and baker, assistant and baker, etc.

Discussion

Basing a robot's behavior on a model of human behavior may appear to be an obvious and straightforward solution to the design of conversational robots. This may be the case for some kinds of behavior, but it was not the case for help-giving communication. Even these preliminary observations raise several concerns about the feasibility of this approach for more socially sophisticated interactions. First, there is substantial variation in how people speak. If we intend to model human behavior, which human are we going to model? One helper says, "Well, uh, if you over beat it, it'll, uh, actually, that won't be good either." Another helper says, "Oh a wee little bit more. Still a little lumpy." Human speakers vary their speech in a number of different and sophisticated ways. We can, of course, rule out obviously undesirable behavior, but that does not go very far in directing which human behavior should be modeled for implementation on a robot.

Identifying appropriate human behavior to model is not obvious. Much of what people do in rich, realistic contexts has ambiguous effects on listeners' impressions. So which behaviors should be modeled? Previous research on human communication behavior, or learning sciences research, can draw our attention to specific ways of speaking, but we cannot be sure that these strategies have the intended effects. For example, one help giver I observed consistently offered listeners' a rationale for the information being offered. One could presume that taking the additional time to explain why the help was being given would be perceived as particularly considerate.

```
EXPERT - DEE:  you didn't chop it small enough. yeah,
               that's why I told you to chop it because
               that way it cooks faster.
```

Alternatively, listeners' could respond negatively to this kind of help-giving behavior because the helper seems to take the tone of a teacher. In the prior example, the bakers did not respond to this baker's utterance at all; they continued with their task in hushed tones to one another. The dynamic between the bakers and this help giver, overall, was not interactive; the helper was most like a supervisor to them. In this situation, the bakers did not sign up for a tutorial on baking, and the additional information could have been unwelcome. These help-giving instances are ambiguous with regards to impression formation. Existing theory may be able to help us conceptualize these interactions, but there are no easy assumptions we can make about the use of this strategy by help givers. This example of providing rationales is one of several verbal help-giving strategies I observed in this small sample of communication that are not easily categorized as desirable or undesirable. A large percentage of language features would be considered ambiguous.

Finally, whether all of the observed communication behavior could be satisfactorily implemented by a conversational robot is highly questionable. One of the leading assumptions in human-robot interaction research is that reactions to robots are much the same as reactions to humans. This idea is so entrenched that sometimes it is possible to forget how relational interaction between people, even strangers in this case, can be. In this sample of conversational behavior, I observed several instances where help givers told stories about their prior baking experiences as a way of offering a suggestion to a baker. For example, Dee tells the bakers they should put water into any empty wells of the cupcake tin. She gives an explanation, "so it doesn't burn," then she offers the additional point, "I always put water in mine."

```
BAKER – ELISE: do you want to fill this with water? the
               empty-

EXPERT - DEE:  yeah, you better. yeah. so it doesn't,
               doesn't burn.

BAKER - ELISE: I don't know why, but that's what my mom-

BAKER - APRIL: -cause the water's going to evaporate,
               right?

BAKER - ELISE: -my mom does

EXPERT - DEE:  I always put water in mine

BAKER - APRIL: really, it doesn't make the cupcakes more
               moist?
```

Dee is the authority here, so her claim that this is what she "always" does carries a bit of weight. This is a very simple strategy for a human help giver, but is this the kind of approach that a robot helper might be able to utilize? What does it mean for a robot to say, "I always put water in mine"? Even supposing the robot was a cupcake-making robot, is this a good strategy for robots? Human help givers referenced their baking experience fairly regularly in my observations. They referred to their prior mistakes. They refer to the period of time when they were still learning. They tell funny stories about baking mishaps. Is this relational approach something that a robot can attempt? Of course, this is empirically verifiable, and it should be explored. Because human behavior is often observed with the explicit goal of engineering appropriate robot behavior, these types of human strategies, the ones that may or may not work for robots, remain unexplored. Understanding these boundary conditions are essential to understanding the full range of human responses to conversational robots.

Baking in a Laboratory Kitchen

In the dormitory, I observed a variety of communication styles that appeared to be related to the age of the helper. Older helpers seemed to be much more direct than peer helpers, for example. In order to observe an even greater variety of communication styles, I wanted to observe different pairings of help givers and help recipients. I needed greater control over the participants in this stage, so I remodeled an experimental laboratory to look like a kitchen. Then, I recruited participants from the Carnegie Mellon community to come by and bake cupcakes together.

Procedure

I recruited participants online, giving a short quiz about baking to determine relative expertise. Participants with low quiz scores were matched up with participants with high quiz scores. Those with low quiz scores were the primary bakers; they were invited to bring a friend to the session too, if they wished. By including two bakers that knew each other, I hoped to observe any possible sensitivity to being seen by a friend while receiving help. Participants with high quiz scores were told they were coming in to make sure the other participants made good cupcakes. Participants in the role of the help giver were not given any specific instructions about how to assist the bakers. They were simply told to give assistance verbally, when necessary, but not to physically make the cupcakes themselves.

Participants were invited to sessions in pairs that created relationships I was interested in observing. Table 4 details the participants in each laboratory baking session. All activity in the kitchen laboratory was recorded with audio and video. After signing consent forms, participants chose one of four cupcake recipes and began working on their cupcakes. The experimenter was not in the kitchen laboratory while participants worked together, except to check on the video camera. When the cupcakes were finished, participants completed a paper questionnaire. The questionnaire asked participants for their impressions of their cupcakes, of baking in general, and of the other participants in their session. The experimenter photographed the cupcakes, and the participants wrapped up their cupcakes to take home with them.

Table 4. The individuals involved in each laboratory baking session.

Session	Helper	Baker #1	Baker #2	Previous Relationships
A	Undergraduate (F)	Undergraduate (M)		Helper and baker are friends
B	Undergraduate (F)	Graduate (M)		No previous contact
C	Staff Member (F)	Undergraduate (M)	Undergraduate (F)	Bakers are friends
D	Staff Member (F)	Undergraduate (F)	Undergraduate (M)	Bakers are friends
E	Graduate (M)	Undergraduate (F)	Undergraduate (F)	Bakers are friends
F	Undergraduate (F)	Graduate (M)	Undergraduate (F)	Helper and baker #2 are friends

Following each session, the conversation that surrounded the preparation of the cupcake batter was transcribed and the transcript was segmented into help-giving instances. A help-giving instance is an utterance or a set of utterances where the helper is offering information, whether because the baker asked a question specifically or because the helper volunteered the information. The number of help-giving instances by a single helper ranged from 23 to 39 during an activity that lasted approximately 30 minutes. A help-giving instance could be a single utterance, or it might include a couple of utterances between the helper and the baker where the pair is clarifying the information the helper provided. These clarification responses are included in further analyses of each help-giving instance. The help-giving instances were iteratively coded for language phenomena used consistently by helpers in their help messages; these phenomena will be described in the next section. These language phenomena were identified by approaching the data with the question, "What linguistic features make these help messages different from the help provided by the cupcake recipes themselves?" This is a question about what makes help-giving communication between people, face-to-face, unique. In further discussion of these analyses, the participant doing the physical baking of the cupcakes will be referred to as the "baker" and the participant verbally assisting the activity will be referred to as the "helper."

Language Features of Help-Giving Communication

As expected, information offered by human helpers deviates from the directive language of a cupcake recipe in several different kinds of ways.

- Agency (I, we, you, implied-you)
- Hedges (I guess, I think, maybe, might, probably, kind of)
- Discourse Markers (you know, yeah, well, so, oh, like, just, I mean, actually, basically)
- Turn Introductions (ok, here, now)
- Fillers (um, uh)

Assumptions of Agency

Helpers sometimes use *I* and *we* as the agent in their conversation instead of the more directive *you* structure. The typical imperative sentence, something like "turn off the mixer," has the pronoun "you" as its subject. This type of sentence can appear domineering, and we observed help givers in this sample avoiding this construction by saying, "I would set it [the oven] for a little bit lower" or "yeah we can go ahead and line the pans too." Although these help givers are not physically involved in this activity their language involves them in the activity as well. All helpers, except for one, used this strategy intermittently in their communication.

Hedges

Hedges limit the truth proposition of a statement (Brown & Levinson, 1987). Prefacing a statement with a hedge like "I think," "I suppose," or "I guess" literally limits the certainty with which the statement is made, but politeness theory suggests that these qualifications are not always interpreted literally. In our sample of helping communication, help givers use hedges even when they appear to be quite confident of their information. In the following interaction, the baker is using a standing mixer to incorporate some flour. The helper thinks the baker should stop mixing the batter. (Cupcakes will develop a denser texture if they are over mixed.)

```
HELPER - JENNIFER:  I think it's probably-

BAKER - GEORGE:     -that's probably enough [turns off
                    mixer]
```

In this analysis, we observed the use of the word "might" and "maybe" as well. Helpers said things like, "it's actually maybe gonna be easier if you come around to where I'm sitting" and "you might, actually you want to get one that's a little bit bigger." In our sample of helping communication there were helpers that used hedges rarely, only two or three times in a session. But there were other helpers that used hedges more frequently, up to twenty times in a session. (Appendix A lists hedges used by helpers in each session.)

Discourse Markers

The class of words known as discourse markers includes words like *you know*, *I mean*, *well*, *just*, *like*, and *yeah*. Discourse markers do not change the meaning of a sentence, instead various research has related discourse markers to interpersonal relationship management, language production, language comprehension, and turn-taking processes (Fox Tree & Schrock, 2002; Schiffrin, 1987). In our sample of help-giving communication, all the helpers use discourse markers consistently, but there was a striking difference in the number of discourse markers used, particularly in a single speaking turn. Appendix A lists each discourse marker used by a helper in an information exchange. Here, an information exchange can be a set of turns between the helper and the baker; it ends with an acknowledgment on the part of the baker. Helpers in sessions A, E, and F use discourse markers more frequently than the other helpers. In the following example, both helpers are offering information about how to fill the cupcake tin with batter. Mary's help message includes a single discourse marker, *just*. Jennifer's message uses *actually* and *yeah*; she also uses *just* and *like* twice for a total of six instances in this speaking turn.

```
HELPER - MARY:       and then when you fill it, you just
                     fill it halfway

HELPER - JENNIFER:   actually yeah I would just like scoop
                     it with the tablespoon and just like
                     put it in until
```

Interestingly, Jennifer does not complete her point because she is interrupted by the baker. It is possible that the punctuation of the message by discourse markers allows greater opportunity for the listener to begin a speaking turn.

Turn Announcements

We also observed helpers using a single word as a turn announcement, words like *here*, *now*, or *ok* that occur only at the beginning of a turn. Several helpers used a turn announcement like this, presumably to get the baker's attention before beginning to offer information. Helpers each had their favorite one and frequently began their help messages with the same word. For several helpers, this word was *ok*.

```
HELPER - MARY:   ok it's in

BAKER - RAQUEL:  oh it's in?

HELPER - MARY:   ok. now what you want to do is alternate,
                 add your egg. ok. then we alternate the
                 milk with the flour
```

Other helpers used *here, now,* or *so* to start help messages. Helpers had different preferences in their choice of turn-opening discourse marker, but they were consistent in their choice.

Fillers

Finally, we observed the presence of filler words like *um* and *uh*. While these words are often considered throw-away words, research has described fillers as indicators of delay or speech production trouble (Brotherton, 1979; Clark & Fox Tree, 2002) and they may have an effect on impressions of the help message. Helpers in this sample did use words like *um* and *uh*.

```
HELPER - SAMANTHA:  here. don't pour in, um, pour like, put
                    the actual scoop into the cocoa and
                    you'll make much less of a mess
```

The use of fillers was quite consistent across helpers with one notable exception; there was a single helper that only uttered a single *uh* during the session.

Summary

The observational stage of this research had two related objectives. First, the observations of naturally occurring help-giving communication were gathered in order to identify specific linguistic features in use in situations like this cupcake baking task. The cupcake baking task represents a class of activity where the desired outcome is open-ended and the necessary steps are somewhat flexible. In this type of situation, I observed help givers using a variety of different linguistic strategies, including hedges, discourse markers, and fillers. In general terms, these linguistic strategies distance helpers from making strong, direct claims on the listeners' subsequent actions, but these observations do not explicitly inspect help recipients' responses to these various strategies. By asking help recipients to view the recordings, it would be possible to ask help recipients themselves to reflect on their own reactions to the helpers with whom they interacted (an interesting direction for future work). In its present form, this analysis is not sufficient to conclude which strategies are desirable and which are undesirable, but this work does document several communication strategies a robot helper could attempt, in socially sensitive situations The second goal of the observational stage of this research was to collect speech from human help givers upon which to base my experimental manipulations in Experiment 4. In Experiments 1, 2 and 3 I generated linguistic manipulations based on prior perspective-taking theory, but the specific word choices and sentence construction were based only on creative intuition about how human helpers might offer information in the tool selection task. The help-giving communication observed in Chapter 4 allowed me to design verbal manipulations based on the manner in which human helpers actually talk.

Chapter 5: Experimental Investigation of Politeness Strategies

The goal of this thesis is to investigate social responses to a robot that offers its listeners help. Chapter 4 reviewed careful observations of human help-giving communication that describe a number of ways that human help givers distance themselves from issuing direct commands. These observed communication phenomena are the basis for the final, experimental phase of this thesis, an experiment that directly compares perceptions of human helping communication with a robot's helping communication. By basing dialogue manipulations on observed language use and by using a human control condition, the design of Experiment 4 seeks to avoid the difficulties in interpretation encountered in the prior experiments (described in Chapter 3). Experiment 4 asks two related questions. First, how do people respond to different help-giving strategies? And, secondly, do responses to human helpers differ from responses to robot helpers?

In Chapter 4, I elaborated on a range of strategies used by help givers to avoid making direct claims or giving direct orders to help recipients. One of the basic approaches in the design of a robot's behavior is to model its behavior on human behavior. Even the most cursory attempt at observing natural human behavior suggests this approach is insufficiently directive. In a socially complex interaction, such as a helping interaction, a number of issues immediately arise when attempting to specify appropriate human behavior upon which to generate appropriate robot behavior. First, which person should be the model? Throughout my observations of help-giving communication, there was substantial linguistic variation across speakers. Some of this variation might be easily excluded as undesirable; just because people speak in a particular way does not mean it is the right thing to do. Even ignoring obviously problematic strategies, there are still a range of help-giving styles and strategies from which to choose.

Second, much of the variation in human communication is difficult to judge as appropriate or inappropriate. I observed that some helpers say *ok* each time they begin a turn. Is this strategy useful, annoying, or neutral? When designing a robot's behavior based on human behavior, existing communication theory will not be able to provide a rationale for every choice. And, third, there are verbal behaviors which do not, at first glance, appear to be appropriate strategies for robot speakers. Human help givers tell stories about their cooking experiences and commiserate over shared mistakes. Whether robots can successfully utilize these strategies is an open question.

The observations of human help givers indicated a range of strategies that may be useful for robot help givers. These strategies provide an opportunity to test social responses to robot help givers by comparing them directly with social responses to human help givers. In this experiment, I wanted to explore the variation in the human help givers' language use so I focused on the use of hedges (words like *I think* and *probably*) and the use of discourse markers (words like *I mean* and *so*). Hedges and discourse markers were used by some human helpers but not others, so they offer an opportunity to test the impact of the linguistic variation observed in human help giving.

There is clear support for the use of hedges as a politeness strategy (Brown & Levinson, 1987), but the support for perceptions of discourse markers are not as clear. I chose to investigate the use of discourse markers because of the ambiguity surrounding impressions of their use. It is true that the consistent use of discourse markers I observed is common practice in a subset of the population, but this does not automatically imply that this behavior generates positive impressions of the speaker. Finally, discourse markers represent a linguistic approach that is not commonly incorporated in our communication with computers or robots. Although using hedges seems particularly appropriate for a robot speaker, the frequent use of discourse markers by a robot could create an uncomfortable impression. Frequent use of discourse markers can be associated with young people (Stubbe & Holmes, 1995) or with difficulty in speech production (Siegel, 2002). It is necessary to evaluate the appropriateness of discourse markers for robot helpers, given the awkwardness inherent in the idea of a robot saying, "Yeah, you know you've just got to like..." In the sections which follow, I describe the theoretical motivations, hypotheses, and experimental design in more detail.

Theory & Hypotheses

In Experiment 4, I investigated the role of hedges and discourse markers in influencing perceptions of help givers. It is important to note that there are different interpretations of the functions of these language features for turn-taking, for repair, and for cohesion. See Schourup for review (1999). Hedges literally express uncertainty; they include qualifying types of language such as *I guess, maybe, probably, I think, sort of*. They are thought to be associated with negative politeness. Negative politeness protects listeners from threats to their autonomy while positive politeness encourages social connection and rapport (Brown & Levinson, 1987). Hedges are a form of negative politeness because they limit the universality of the statement, allowing the listener to disagree if necessary. Hedges are not always interpreted as real qualifiers of the content of the message. It seems hedges are a conventional way that speakers mitigate the force of their communication. On the basis of politeness theory, I hypothesized that speakers using hedges in their help messages will be perceived as more considerate and less controlling.

H1: Speakers using hedges will be rated more considerate and less controlling than speakers not using hedges.

The second linguistic strategy investigated in this experiment is the impact of discourse markers. The role of discourse markers in influencing perceptions of speakers is disputed. Discourse markers include words like *you know*, *I mean*, *well*, *just*, *like*, and *yeah*. These words operate at a pragmatic level; their meaning is derived not exclusively from their literal definition but from their use in context. From observation of human help-giving communication, I focused particular attention not to these words in isolation but to their consistent, repetitive use. It is the combination of multiple discourse markers used in close proximity that may particularly impact perceptions of the speaker. There are several different interpretations of the use of discourse markers. See Fischer for a review of several theoretical approaches to the current understanding of discourse markers (2006a). First, discourse markers are thought to be related to speech production difficulties. So one interpretation of speakers who use *like* frequently is that they are having trouble finding the right words (Siegel, 2002). The phrase *you know* also may indicate the speaker is stalling for time (Holmes, 1986).

Second, the use of discourse markers is associated with more casual speech or speech between young people. In particular, the phrase *you know* has been found to be more prevalent in the speech of young people (Stubbe & Holmes, 1995). *You know* invites addressees to make an unspoken inference (Jucker & Smith, 1998). It may be common within younger communities or in more casual situations because speakers in these contexts are more willing to engage with their addressees' interpretations (Fox Tree & Schrock, 2002). Between young people, discourse markers may appeal to positive politeness by reinforcing the similarity between the speaker and the listener.

Finally, discourse markers are associated with negative politeness. By saying *you know* but remaining vague about the details, speakers may be engaging in a form of negative politeness. Using *you know* allows helpers to be less explicit about the direction they are giving and to give greater weight to the listeners' interpretations. Andersen describes talk involving the use of *like* as "loose," meaning it reduces the commitment of the speaker to what follows (Andersen, 2000). *I mean* warns of upcoming adjustments (Schiffrin, 1987). Helpers who frequently use *I mean* in their help messages may be more comfortable adjusting their messages as they are being produced, or they may be particularly sensitive to the way the message could be interpreted (Fox Tree & Schrock, 2002). The use of *I mean* may signal a casual, flexible way of speaking that could impact negative politeness as well. If the way a message is given is in flux, then it may signify the speaker is not as attached to it and would modify if pressed.

There is no single theory that predicts perceptions of discourse marker use, especially as is the case with this experiment, when I am exploring the use of discourse markers as a style of speech that includes the repetitive use of several different discourse markers simultaneously. Nevertheless, the prior work on individual discourse markers suggests that discourse markers can distance the speaker from making a strong, direct claim on the listener, therefore I hypothesized that the use of discourse markers would improve perceptions of the speaker as more considerate and less controlling.

H2: Speakers using a style of speech containing several discourse markers will be rated more considerate and less controlling than speakers not using discourse markers.

The first two hypotheses anticipate a benefit for the use of both hedges and discourse markers. Although there is no specific prior work that observes the combination of these two communication strategies, it is possible that their combined effect would be additive. That is, the use of both strategies together could be better than either strategy alone. In the observational work, previously described, the use of frequent discourse markers was accompanied by the use of hedges, so I hypothesized that their combination would be particularly effective.

H3: Speakers using both hedges and discourse markers will be rated more considerate and less controlling than speakers using either strategy in isolation.

As described thus far, my hypotheses for these experiments rely on the human-human communication literature and apply to perceptions of human speakers only. In addition to testing perceptions of human speakers using hedges and discourse markers, this experiment also seeks to test perceptions of robot speakers. The prior work in human-robot communication strongly favors the hypothesis that perceptions of and interactions with robots mimic interactions with humans (S. Lee, et al., 2005; Powers, et al., 2005; Reeves & Nass, 1996). Based on the prior work on social responses to technology, we would not anticipate significant differences between human helpers and robot helpers, but the communication strategies being tested in this experiment, particularly the use of discourse markers, are thought to be associated with spoken human speech and may not create the desired effect when used by robot helpers. The frequent, consistent use of discourse markers, the style being tested in this experiment, may seem unnecessary and affected when spoken by a robot. For this reason, I predict that discourse markers will be more effective when used by human helpers than when used by robot helpers.

H4. Human helpers using discourse markers will be rated more considerate and less controlling than robot helpers using discourse markers.

Method

In order to compare impressions of speakers using hedges and discourse markers, I used a within-subjects design (2X2, hedges X discourse markers) to expose participants to each of the linguistic strategies. A direct help message (with neither hedges or discourse markers) was created. Three additional help messages added hedges, discourse markers, and their combination, for a total of four linguistic, help-giving styles. I based these help messages on the steps of the cupcake baking task where participants struggled during the laboratory cupcake baking sessions.

If the manipulation included hedges, the help message used a single hedge. In my prior observations, hedges were used consistently by certain helpers when they offered information, but they were used one at a time. If the manipulation included discourse markers, the help message included three to four discourse markers. This manipulation specifically tests the combination of frequently observed discourse markers. from the observational phase of this thesis. For each of these four linguistic conditions I created a script for a short, video vignette involving a baker and a helper. Table 5 details the help messages in the scripts for each linguistic condition.

To compare impressions of human and robot helpers, I produced these video vignettes with a human helper and then digitally spliced a robot helper over the human helper in each video segment. Video vignettes were produced for all four linguistic conditions, one set showed a human speaker and a second set showed a robot speaker. Participants viewed all four linguistic conditions. Two conditions were shown with human speakers, and two conditions were shown with robot speakers. Ordering was counterbalanced. After viewing each video vignette, participants responded to a questionnaire.

Table 5. Experiment 4, help messages communicated in each linguistic condition.

	hedges & discourse markers	**no hedges & discourse markers**	**hedges & no discourse markers**	**no hedges & no discourse markers**
measure flour	So I think you want to use a spoon to measure the flour into the cup. Cause like you don't want to pack it. Yeah, scooping with the measuring cup packs the flour.	So yeah, you don't want to pack the flour by like scooping it out. Fill the measuring cup using a spoon.	When you measure it, I think you should put the flour into the measuring cup with a spoon. That way the flour's not packed in there.	Scooping it out with the measuring cup packs the flour. You don't want the flour to pack. Put it in with a spoon instead.

	hedges & discourse markers	no hedges & discourse markers	hedges & no discourse markers	no hedges & no discourse markers
cream butter and sugar	And kind of mix it, until it's just like fluffy. Basically, a nice smooth consistency, a little bit lighter color.	Basically just keep going until it's like a smooth mixture. Lighter color and fluffier.	Until the batter looks smooth. It'll get kind of fluffier and the color will lighten up.	The mixture should be smooth and fluffy. The color will get lighter too.
scrape the sides of the bowl	You probably want to just scrape down the sides. Like with a spatula, yeah. I mean, get everything mixed in.	Like scrape around the sides with a spatula. I mean, yeah, just mix all that stuff in there.	You can probably scrape the bowl. You want to use a spatula to get everything on the sides mixed in.	Scrape down the sides of the bowl. With a spatula. Mix the stuff on the sides into the rest.
pour batter	Oh, so actually it would maybe work better if you used a small measuring cup, or an ice cream scoop to fill those.	Oh, so you can use a small cup or actually an ice cream scoop to make it easier.	Maybe drop it in there with a measuring cup or that ice cream scoop. It'll work better.	Use a small measuring cup or an ice cream scoop. It's easier than a spoon.
fill cupcake tins	I think you can fill them more. So leave like a little room. Three-quarters full, you know.	You know, filling them more, three-quarters full, will be ok. So like add more batter to each of these.	Keep going, I think. You can add more batter. Three-quarters full is good.	Add a little more batter to each one. Fill them up, three-quarters full.

Participants

Seventy-seven Carnegie Mellon University students and staff members, as well as members of the general Pittsburgh community, were recruited from Carnegie

Mellon's experiment scheduling website. They were each paid $10 for their participation in this experiment.

Procedure

This experiment was conducted in a laboratory where the videos could be projected on a large screen with accompanying high-quality audio. The experiment was conducted in small groups of between three and six participants at a time. When participants arrived at the laboratory, they were asked to answer several questions about baking, as a test of their experience with the task domain. These questions illustrated a step of the baking process with a picture and asked participants to decide if there was any correction they would make to the baker's process. If they indicated they would correct this baking step, they were asked to write in what they would tell the baker if they were offering the baker help in this sceanrio. Participants were then informed they would be watching a series of short videos and would be asked to answer several questions following each video. Questions from the participants about the experimental paradigm were very rare. The experimenter sat at the back of the room to turn the videos on and off and was not visible to the participants during the session. Four video clips were shown. After each video clip, participants were asked to respond to several pages of items on a questionnaire. After viewing the final video clip, participants were paid for their participation and invited to ask any remaining questions they had about the procedure or the experiment.

Because the experimental procedure requires counterbalancing along three different dimensions (linguistic condition, human or robot helper, and actor), it was not feasible to take an exhaustive counterbalancing approach. Instead, I conducted a total of sixteen sessions during which each linguistic condition was counterbalanced for the order of presentation, and the actor in each video clip was counterbalanced for order of appearance and for linguistic condition used. Figure 6 shows two examples of the way this counterbalancing approach generated four videos in each session (one for each linguistic condition), showing two robot helpers and two human helpers, and using each actor only one time.

Table 6. Two examples of how counterbalancing was conducted across sessions.

	1st Video	2nd Video	3rd Video	4th Video
Order A	Human	Robot	Human	Robot
	Max	Ada	Kit	Rob
	Hedges & No Discourse Markers	No Hedges & No Discourse Markers	Hedges & Discourse Markers	No Hedges & No Discourse Markers
Order B	Robot	Human	Robot	Human
	Kit	Rob	Max	Ada
	Hedges & Discourse Markers	No Hedges & Discourse Markers	No Hedges & No Discourse Markers	Hedges & No Discourse Markers
Counterbalancing continues in this way for a total of 16 unique presentation orders...				

Video Stimuli

Four scripts were created for use in the video vignettes, based on the four linguistic conditions previously described. Each script contained help messages for five steps in the cupcake baking task from measuring the flour to filling the cupcake tin with batter. The steps were identified as sources of frequent help giving in the prior observational work. Each video vignette was approximately three minutes in length and included a short introduction to the characters and the task. Following this introductory material, a statement of the current step and the baker's response to it was displayed before showing the help giving interaction. For example, prior to seeing the helper offer information about how long to mix the butter and sugar together with an electric mixer, participants saw the following message displayed on the screen: Evan reads from the recipe, "'Cream butter and sugar." In the subsequent frame, Evan is shown mixing the butter and sugar. Then, the helper is shown speaking the help message. Over the course of the three-minute video clip, five different baking steps are introduced. After each step is introduced, the help message is delivered.

Because different speakers using the exact same language (hedges and/or discourse markers) may generate different impressions based on paralinguistic qualities or physical appearance, I captured the dialogue from each script using four different actors. The actors were referred to in the videos with pseudonyms (Ada, Max, Rob, and Kit). I used two men and two women in order to separate the influence of individual characteristics from the verbal aspects of the help messages. To do this, I hired four undergraduate drama majors to act out all four scripts, two men and two women (see Figure 12 for still images of each actor in the laboratory kitchen). By counterbalancing which actor was seen in which linguistic condition, it is possible to investigate the effect of the linguistic condition independent of the effect of the actor's appearance, voice characteristics, or other potentially influential factors.

Figure 12. Two men and two women acted the part of the helper for each script.

I created a version of each linguistic condition using four different actors, so in total there are sixteen video clips showing human help givers. To compare impressions of robot helpers, I duplicated the set of video clips showing human help givers and overlaid video of a robot helper directly on top of the human help giver. In this way, the baker's behavior is kept consistent and the image of the helper is the only visual difference between the sets of videos (see Figure 13 for still images of the videos with the robot helpers digitally inserted into the frame). The robot is designed with variable forehead and chin shapes. These modifications were used to create four robots with slightly different facial features, to stand in for the four human actors in the videos. These robots were identified in the videos with the same pseudonyms as the human actors (Ada, Max, Rob, and Kit).

Figure 13. Video of a robot helper was laid over the human helper in the previously created videos.

Because of the acoustic variability in the way the human helpers spoke (every actor spoke their lines in a subtly different way), it was important to keep those paralinguistic features consistent in the videos containing a robot helper. Synthetic speech is typically used for a robot's communication, but it was difficult to recreate the subtleties of naturally occurring speech (particularly speech containing discourse markers) with speech synthesis. Using the recorded human speech as a base, I modified the audio track to create a more metallic, harmonic sound. The audio track was modified by duplicating the track, changing the pitch on the duplicate track, and lowering the volume on this second track. This audio modification was done to create the impression that the audio in the clip was the robot's voice.

Measures

I observed participants' reactions to the help-giving scenarios with both quantitative and qualitative measures. After viewing each video, participants were asked to rate their agreement with a series of statements about the helper, the help message, and the task itself. After considering their agreement with each statement on the questionnaire, participants were asked to describe the video, in their own words, for someone who had not seen it.

To reduce the number of dependent variables, I conducted a factor analysis on the fifteen questionnaire items that ask participants for their perception of the helper and the helper's communication behavior (see Table 7). I conducted a principal components analysis with a varimax rotation. Using an eigenvalue threshold of 1, the analysis revealed 4 factors accounting for 78% of the variance. The first factor describes participants' impression of the speaker as a considerate and supportive helper. I labeled this factor "Considerate." The scale is reliable (Cronbach's α = 0.86). The second factor describes participants' impression of the speaker as controlling. This scale is more reliable without the third item, "knows a lot about baking" so the "Controlling" scale was used with only the stronger two items (Cronbach's α = 0.79). The final factor in the analysis describes various aspects of the helper's fluency. These items did not achieve reliability as a scale, so this third scale was not included in later analyses.

An additional scale was used to measure participants' reactions to the videos. Participants were asked to rate their agreement with statements such as, "I like the helper" and "If I were making cupcakes, I would want this helper around." These statements indicate positive rapport with the helper. Three related statements were included in this "Liking" scale; as a scale, these three items were highly reliable (Cronbach's α = 0.86).

Table 7. Factor analysis of questionnaire item loadings

	I	II	III	IV
The helper is tuned into the baker's needs.	0.81	-0.08	-0.02	0.02
The helper talked to the baker like they're friends.	0.80	0.09	-0.04	-0.004
The helper was not considering the baker's feelings.	-0.77	0.24	0.20	0.03
The helper said some rude things to the baker.	-0.71	0.24	0.19	0.02
The helper is not supportive of the baker.	-0.70	0.28	0.14	0.14
The way the helper offered help was diplomatic.	0.65	0.12	0.22	0.31
The helper wants to show the baker approval.	0.63	0.12	0.26	0.0006
The helper took control of the baking activity.	-0.33	0.72	-0.18	-0.05
The helper spoke strongly about how the cupcakes should be made.	-0.33	0.68	-0.31	-0.02
The helper knows a lot about baking.	0.31	0.67	0.01	0.09
The helper hesitated while offering advice.	0.11	-0.14	0.75	0.11
Offering help is effortless for the helper.	0.11	0.07	-0.62	0.43
The helper speaks fluidly.	0.15	0.36	-0.43	-0.13
The helper offered suggestions that the baker can either accept or reject.	0.49	-0.27	-0.01	0.54
People like the baker don't talk the way the helper does.	-0.39	0.07	0.04	0.70

In addition to the ratings of agreement collected from participants, I also asked them to describe, in their own words, their interpretation of what the helper was doing in each of the videos. These open-ended responses were transcribed and coded according to the Linguistic Category Model (Coenen, Hedebouw, & Semin, 2006). The Linguistic Category Model attempts to measure the level of abstraction with which people describe individuals. I used this coding manual to explore the differences between the way that human helpers and robot helpers were described in open-ended responses. The coding manual defines five different codes: Adjectives, Descriptive Action Verbs, Interpretative Action Verbs, State Action Verbs, and State Verbs. Instances of the various verb categories were rare in this sample, to describe either human or robot helpers. The Adjective category was more common. Descriptions in this category, for example "the robot was aggressive," suggest that participants are ascribing qualities or traits to the helper, such as being"honest" or "aggressive." By using the adjective form, participants are describing helpers as having these qualities indefinitely, as opposed to engaging in a one-time activity, for example "the robot offered information about baking cupcakes." Analyses of the use of adjectives to describe helpers will be discussed more completely in the related results section.

Results

First, I considered the effects of linguistic condition and the impact of the human/robot helper variable on the set of questionnaire scales identified in the factor analysis: Helper is Considerate, Helper is Controlling, and Liking for Helper. Our model includes the effect of hedges, discourse markers, human/robot helper, and all related two-way and three-way interaction effects between those primary variables of interest. The model also includes several influential control variables: order seen (1-4), helper gender, participant age, number of years in the US, baking expertise (number correct from pre-test). Variables such as gender, age, and length of time living in the US were gathered to investigate possible interactions between participant characteristics and perceptions of the linguistic manipulations. It is possible, for example, that second-language English speakers would have different reactions to American politeness strategies. To remain open to these considerations, I included these variables in the analysis.

The use of hedges and discourse markers did increase participants' agreement that the helper was considerate. Hedges and discourse markers alone, as well as both together, improved ratings of considerateness over direct speech, F (1, 211) = 23.49, $p < .001$.(See Figure 14.) Whether the participant saw a human or a robot helper did not have any significant effect, but male helpers were rated as significantly more considerate as compared to female helpers, F (1, 212) = 22.96, $p < .001$.

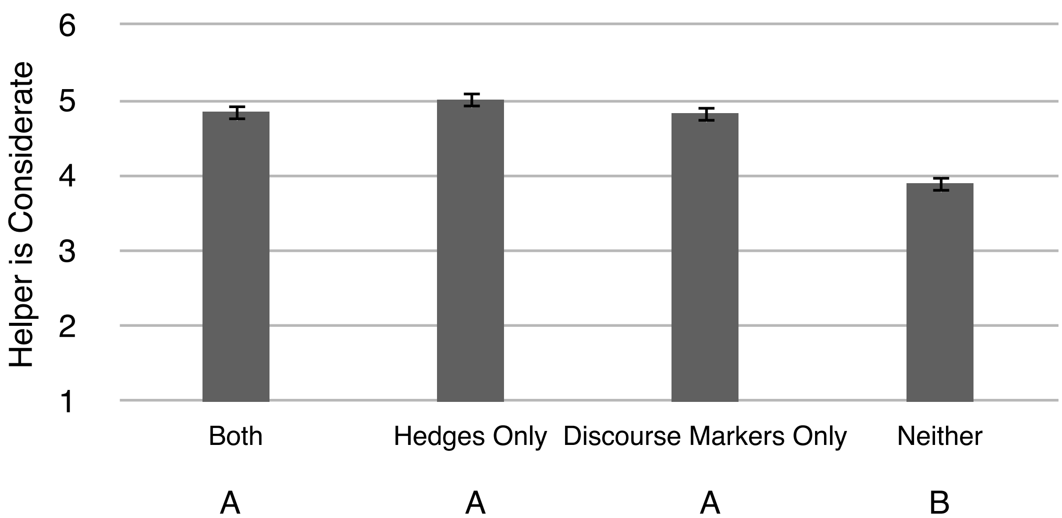

Figure 14. Experiment 4, the use of both hedges and discourse markers significantly increased perceptions that the helper was considerate, but the combination did not improve perceptions any further.

The other questionnaire scale related to perceptions of the helper's language use was the extent to which participants found the helper to be controlling. Again, there was a significant interaction, F (1, 209) = 4.9, $p < .05$, indicating a significant reduction in ratings of the helper's control when either hedges, discourse markers , or both were used. (See Figure 15.)Again, I found that male helpers were perceived to be less controlling than the female helpers even when the words they are using are identical, F (1, 210) = 5.5, $p < .05$.

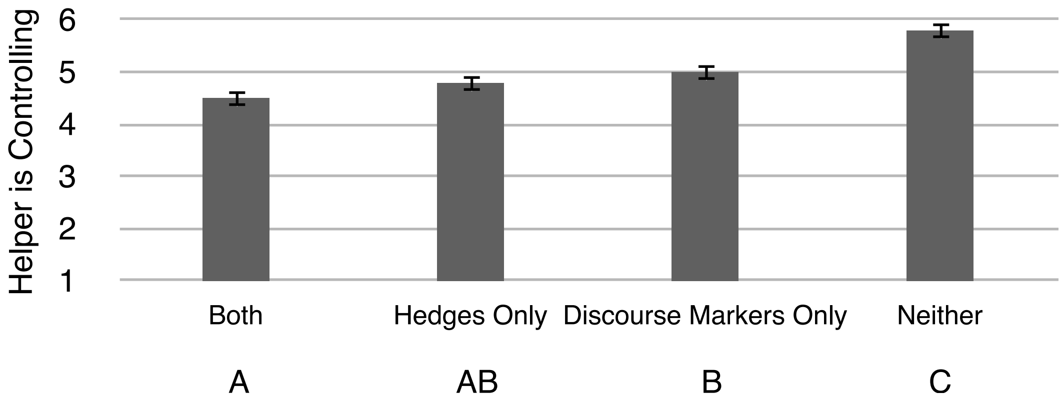

Figure 15. The use of hedges and discourse markers both significantly decrease ratings of control.

Perceptions of control were significantly affected by whether the helper was a human or a robot. A main effect for the human/robot variable indicated that, overall, the robot helpers were perceived to be less controlling than the human helpers, $F\ (1, 209) = 6.2$, $p < .05$. A significant interaction indicated that the impact of the robot helper on ratings of control was particularly evident when the robot used discourse markers, $F\ (1, 277) = 5.28$, $p < .05$. (See Figure 16.) A robot helper using discourse markers was rated as significantly less controlling than a human helper using the exact same help messages.

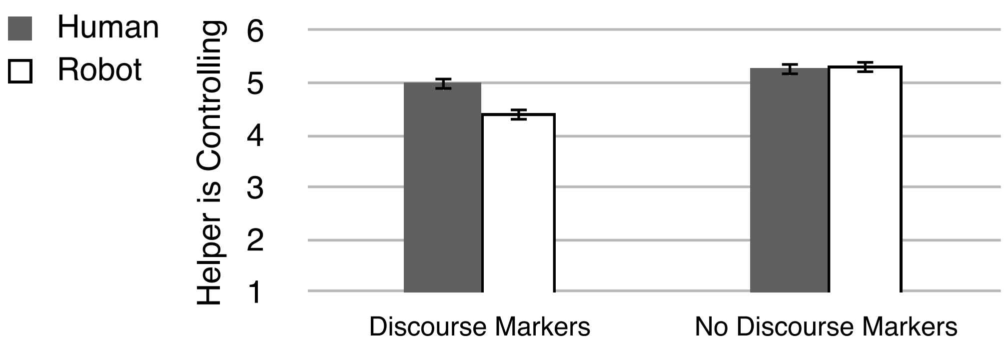

Figure 16. The use of discourse markers is perceived to be less controlling, even more so when discourse markers are used by a robot.

The third questionnaire scale used to measure impressions of helpers in this experiment was a general measure of liking for the helper. The results followed a similar pattern as judgments of whether the helper was perceived as considerate and controlling. When the helper used hedges, discourse markers, or both, participants liked the helper more than when the helper used direct speech, F (1, 210) = 7.73, $p < .01$. (See Figure 17.)

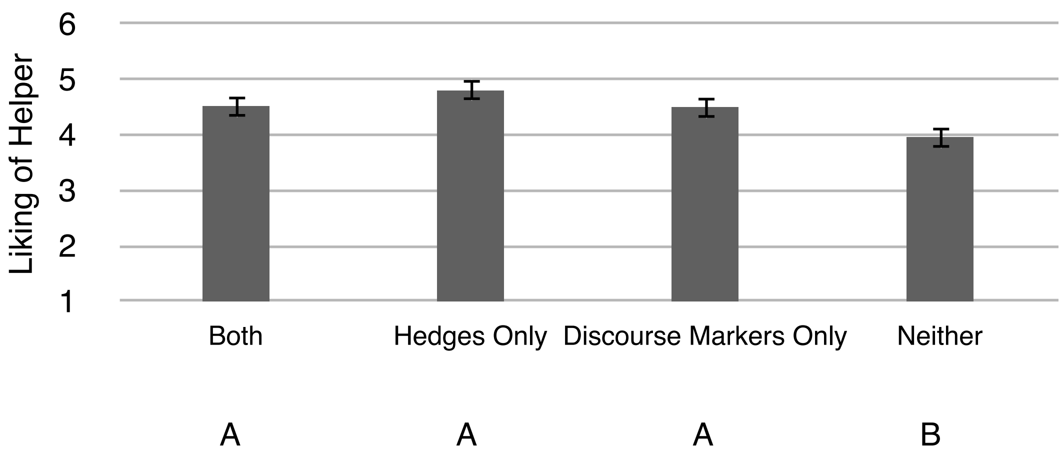

Figure 17. The use of hedges, discourse markers, or their combination all increase measures of liking equally.

Several of the control variables influence the measure of liking as well. Male helpers were better liked than female helpers, F (1, 211) = 17.6, $p < .001$. The older the participant, the more the participant liked the helpers overall, F (1, 76) = 5.97, $p < .01$. The longer the participant has been a resident in the United States, the less the participant liked the helpers overall, F (1, 73) = 4.5, $p < .05$. And, finally, there was a significant interaction between participant expertise and the use of the hedging strategy. The less expertise a participant had, the more the participant liked helpers that used the hedging strategy.

To summarize thus far, I will compare the results of the questionnaire measures with the predicted hypotheses. As predicted in H1, helpers using hedges were better liked, perceived to be more considerate and less controlling. As a communication strategy, hedges and discourse markers were successful for both human and robot helpers. As predicted by H2, helpers using discourse markers were also better liked, more considerate, and less controlling. Discourse markers, it seems, are also a successful strategy. But surprisingly, these two strategies were not more effective when they were combined. This evidence does not support H3, the hypothesis that the combination of hedges and discourse markers would be more successful than either one in isolation. Both strategies made relatively little impact when the other strategy was already present.

Next, consider the intended comparison between human and robot helpers. There were no differences between perceptions of human helpers and robot helpers when the help messages were direct (and used neither hedging nor discourse marker strategies). Both humans and robots were perceived less positively when making direct claims on the help recipient. When both robot helpers and human helpers used hedges, impressions were improved, but when robot helpers used discourse markers they were perceived as even less controlling than human helpers. In H4, I predicted that discourse markers would be more successful for human helpers than it would be for robot helpers. So there is limited support for the idea that robot helpers are perceived differently than human helpers, but in the opposite direction from the way I hypothesized.

In addition to the questionnaire measures, I also collected open-ended descriptions of the videos from the participants. These descriptions were written after each video was viewed, in order to explore whether participants described robot and human helpers differently. I measured this by coding each description for the use of adjectives describing the helper. The use of an adjective, such as "the aggressive helper," indicates the participant is ascribing a quality or trait to the helper, as opposed to saying something like, "the helper acted aggressively" which indicates the helper acted this way in a single instance.

I found more frequent use of adjectives when participants were describing human helpers than when they were describing robot helpers. (See Figure 18.) Using a log transformation to normalize the distribution of adjectives across participants and controlling for total number of words in each participants' description, I found the use of adjectives to be significantly more frequent in descriptions of human helpers than in descriptions of robot helpers, F (1, 218) = 8.1, $p < .01$. I did not find significant evidence that the use of adjectives to describe robot helpers affected participants' ratings of the helper as considerate, controlling, or likable.

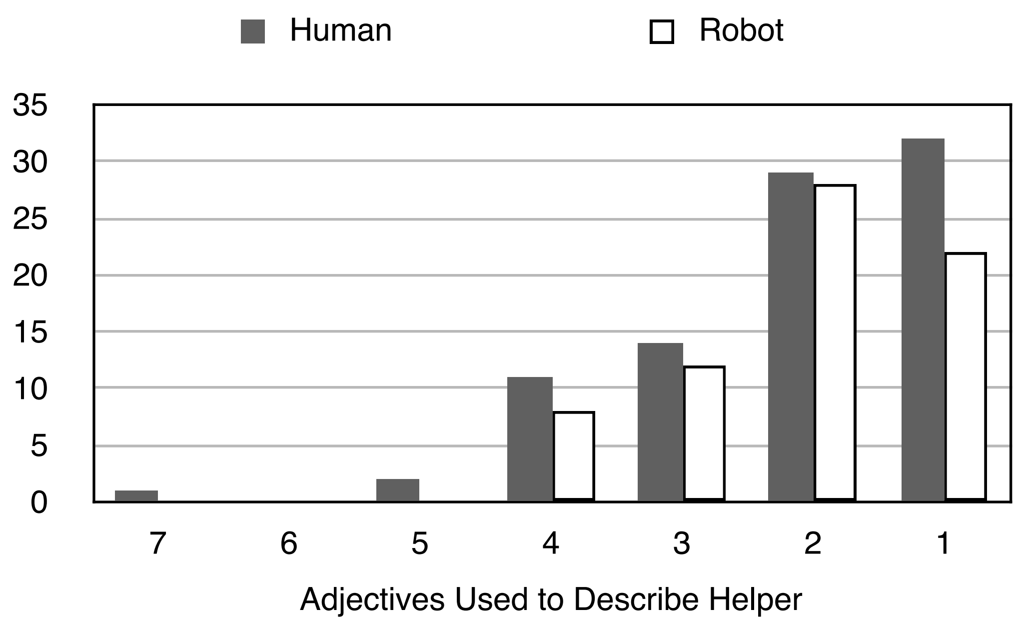

Figure 18. The frequency histogram of adjectives used by participants to describe the helpers.

Limitations

This experiment was designed to compare human helpers and robot helpers in as controlled a manner as possible, but there are a number of trade-offs associated with this design. First, this experiment uses a video vignette paradigm that does not require the participants to interact directly with the human or robot helper. This could influence the participant to take the helper's communication less seriously, as the communication has no instrumental impact on their activity; they are merely bystanders. This could result in ratings that are milder than they might be if the consequences of the help messages were more salient for participants. On the other hand, when participants are viewing the scene as bystanders they have greater opportunity to notice and to dwell on the differences between the help givers, something they might not have the opportunity to do if they were engaged in baking the cupcakes themselves. This intense concentration on the help giver's communication could have accentuated participants' responses to the different strategies.

Secondly, this experiment used actors to reenact scenes that were previously observed in naturally occurring speech. Although the actors did an excellent job with their lines, the dramatic characterization of the interaction may have unforeseen effects.

Third, the particular instantiation of the help messages, with the use of a single hedge and three or more discourse markers, may not allow us to generalize to other uses of hedges or other uses of discourse markers. I observed this pattern of communication behavior in naturally occurring speech, but it is not possible to claim the impressions found in this investigation would hold if the frequency and combination of hedges and discourse markers were varied in another way.

Fourth, this experiment varied the use of hedges and the use of discourse markers but did not combine this language use with other potentially influential factors, such as nonverbal communication behavior, including the physical distance between the helper and the baker, as well as prosodic cues. It is possible that the use of hedges or discourse markers could be quite different in a different tone of voice or a different speed of speech. This experiment used four speakers to vary some of these aspects of communication, but I cannot exclude the possibility that other nonverbal communicative elements, in combination with the linguistic manipulations, would alter impressions.

Fifth, impressions of helpers and of help messages are undoubtedly influenced by factors outside of the scenario depicted in these video vignettes. Help recipients care more about some activities than others, and this may sensitive them to help-giving interactions around activities they strongly identify with (as described in Chapter 1). Various additional elements of the task, the help recipient, and the help provider could, undoubtedly, influence responses to receiving aid. For the sake of simplicity, this experiment looks closely at the influence of the help message, but this influence may weaken when the impact of other situational elements are factored into the equation.

Finally, this exploration of linguistic mitigation strategies is culturally limited to American English. The use of discourse markers, as a phenomenon associated with young, casual speech, could even be limited geographically to specific regions of the United States. With future research, we may be better able to reflect on the generalizability of positive impressions of discourse markers, but for the present, I acknowledge the limitations of this exploration of politeness as culturally bound.

Discussion

The goal of this experiment was to explore the quality of social responses to robots by comparing them directly to social responses to people. The verbal manipulations used in this experiment were based on observed human dialogue behavior. In the previous stage of this research, I identified a number of strategies that are likely to impact social impressions of a help giver. I isolated two strategies for the purposes of this experiment, hedges and discourse markers. I tested the impact of these strategies, both individually and in combination, on impressions of human and robot helpers by asking participants for their reactions to a series of video vignettes.

Participants responded positively on several measures to the use of hedges, as well as the use of discourse markers, but the combination of the two strategies was not any more successful than the individual strategies alone. This is somewhat surprising, given the consistency with which I observed the combination of hedges and discourse markers in naturally occurring speech. Based on these observations, I anticipated that the combination of the two strategies would be the most successful approach. Using the judgments obtained in this experiment, however, I could not determine any benefit for using more than one strategy at a time.

Reactions to robots were, for the most part, similar to responses to humans. Direct speech, without either mitigation strategy, was judged to be the least desirable on several dimensions for both humans and robots. This means that robots will be held accountable for communication that might appear to be rude, if made by a human. Developers of robots that give directions, offer information, or provide other kinds of help cannot assume that a direct approach will be forgiven because it is used by a robot speaker. Robots need to consider the emotional context of their speech and mitigate their communication with politeness strategies.

One strategy, in particular, appears to be particularly successful for robots. Surprisingly, the use of discourse markers improved perceptions of robots more than it improved perceptions of people. For example, when a robot says "I mean, yeah, just mix all that stuff in there," the robot is perceived to be significantly less controlling than when a human helper says the same thing. This result was unanticipated, given the unfamiliar nature of this kind of casual speech on a robotic platform, but it is possible that this lack of familiarity is the source of this effect. Expectancy violation theory was first identified with respect to nonverbal behavior but has subsequently been explored for verbal behavior as well (Burgoon, 1978; Burgoon, Le Poire, & Rosenthal, 1995). When something is unexpected, expectancy violation theory predicts greater levels of arousal and a subsequent stronger reaction to the stimuli. In this case, the use of discourse markers has a positive impact on perceptions, so it has even a stronger effect when it is unexpected, as it is when used by a robot.

This is one of the first experiments to encounter differences between the pattern of social responses to people and the pattern of social responses to robots. This experiment did not find that human linguistic strategies were ineffective for use by robots, but, quite the opposite, this experiment found that human linguistic strategies were even more effective when used by robots. Regardless of linguistic condition, participants viewed robot helpers as being significantly less controlling than human helpers overall. When the robot used discourse markers, in particular, robots were seen as less controlling. This result deserves further investigation. The analysis of participants' descriptions of the interaction in the videos is also suggestive that people may have subtly different conceptualizations of robots. This study found that participants used more adjectives in their descriptions of humans overall, than in their descriptions of robots. Technology designers are typically discouraged from devising interactions that surprise people or are not what they expect, but this lack of similarity between humans and robots may be a positive opportunity to defy expectations, resulting in less domineering robot helpers.

Chapter 6: How Robots Can Help

The goal of this thesis is to explore impressions of a robot that gives help, paying particular attention to how robots might support help recipients' social and emotional well being. Social intelligence is widely acknowledged to be a necessary component of robot helpers, but the pursuit of social intelligence has been interpreted to mean that robots should mimic people. The application of this approach to the design of robot helpers is not straightforward, in socially sensitive situations such as help giving, mimicking human help givers does not necessarily produce positive outcomes for help recipients. In order to clarify how human communication behavior can be used in the design of robot helpers, this thesis has described several human help-giving strategies and compared the impact of these strategies when used by human and by robot helpers.

First, this thesis asks the question: How do people give help? In Chapter 4, I described direct observations of human help-giving communication, in the field and in the laboratory, undertaken to document a range of linguistic strategies used by help givers to distance themselves from making claims on their listeners. These observations highlight the difficulty in modeling human communication behavior for the design of robots. There is substantial variation in the ways people offer help. Which of these styles is appropriate? Who is the appropriate model for a robot's communication?

Second, this thesis tests these help-giving strategies to determine what effects they have on their listeners and asks the question: How do help recipients respond? Just because a human help giver chose to communicate in a particular way does not mean it is the right approach or that it will generate positive outcomes for help recipients. To investigate which conversational features are truly helpful, in Chapter 5 I described the results of an experiment investigating reactions to help-giving communication using hedges and discourse markers. I based my linguistic manipulations on observed human speech, and I directly compared perceptions of human helpers with perceptions of robot helpers.

Third, this experiment allows me to address the question: Is there a difference between human and robot helpers? This experiment demonstrates an alignment between perceptions of human and robot speakers on the negative effects of direct help messages; these messages are perceived to be more controlling. There is some indication that polite strategies such as discourse markers have a stronger positive impact for robot speakers than for human speakers. While not typically recognized as a politeness strategy, the use of discourse markers improved perceptions of robot helpers, in much the same way as hedges.

Theoretical Contributions

Politeness theory is the primary theoretical basis for the identification of hedges and discourse markers as responses to face-threatening situations (Brown & Levinson, 1987). Hedges are one of the most common, and frequently studied, linguistic elements of politeness. Experiment 4 supports the theory's claim that the use of hedges generates a polite impression. Further, speakers using hedges were not perceived to be unsure, inarticulate, or inexpert. The use of hedges increased the perception that the help giver was considerate.

The findings in Experiment 4 related to discourse markers were more surprising. Discourse markers are not included in the catalog of politeness strategies articulated by Brown and Levinson. There are conditions under which one might conclude that discourse markers distance speakers from making a strong commitment to their statement, therefore functioning as a politeness marker. Experiment 4 offers evidence that indeed discourse markers are perceived, much like hedges, to be a politeness strategy. Perhaps even more surprising, Experiment 4 found that the combination of hedges and discourse markers did not have any additional improvement on perceptions of speakers. Politeness theory describes thirty different linguistic strategies that indicate politeness. The variety of strategies is vast, and these strategies commonly co-occur in naturally occurring speech. The results of Experiment 4 show that using hedges and discourse markers in the same utterance does not achieve any additional benefit over using one of them alone. Further work is needed to explore the effect of combining politeness strategies and to investigate how politeness strategies are appropriately and inappropriately combined. Is a combination of multiple politeness strategies even noticed by the listener? Or are listeners more inclined to notice only when there are no politeness strategies at all? To advance politeness theory, we need a greater understanding, not of individual linguistic features, but of how these features function together in various combinations to form polite impressions.

The primary theoretical basis for the comparison between human helpers and robot helpers is the Computers as Social Actors Theory (Reeves & Nass, 1996). This theory describes social responses to technology as adhering to the same basic patterns observed in social responses to people. In contrast to my expectations, the results of Experiment 4 support the CASA approach. Robot helpers using discourse markers made a more positive impression than robot helpers that did not use discourse markers. This effect is even stronger for robots than it is for human helpers. This finding supports CASA predictions, but it is interesting to consider social responses to technology as different, even stronger, than social responses to people. Social responses to robots may be more simplistic and stereotyped. For certain situations, this reaction can be used to generate positive outcomes. A robot using discourse markers is considered friendly, for example, whereas a teenager might be criticized as inarticulate for saying *like* too many times. Across linguistic conditions, robot helpers were perceived to be less controlling than human helpers. If robot developers could find ways to further exploit this difference, the potential exists to create a robot helper with even more sensitivity to the help recipients' independence. Rather than causing discomfort, differences between perceptions of humans and robots may be advantageous.

Because a robot using discourse markers may have violated participants' expectations, one interpretation of this difference between robot helpers and human helpers comes from the literature on Expectancy Violation Theory (Burgoon, 1978; Burgoon, et al., 1995). Perceptions that the robot is being less controlling when using discourse markers is a positive version of expectancy violation; the robot communicated more sensitively than was anticipated, therefore the effect was heightened. Robot developers and designers generally mimic human behavior because that behavior will be familiar and interpretable to the people with whom the robot interacts. With the finding in Experiment 4, expectancy violation theory suggests a different opportunity, that robots might be able to surprise listeners with their behavior and improve outcomes even further. An obvious criticism of this approach would be that long-term interaction might equalize the impact of this effect, but it is equally likely that a strong, initial response to a robot might frame later interactions, over time, even if the robot's behavior ceased to be surprising. Further work is necessary, of course, to understand how an expectancy violation effect might be generated beneficially, but this thesis offers one example of its practical application and benefit.

Guidance for the Design of Robot Helpers

One of the motivations for this work was to explore a situation in which social responses to robots was not altogether positive, where the perception that the robot was a machine rather than a person could be advantageous. Since the robot cannot physically take control of the task because it has no arms, I considered the possibility that helpers would not feel threatened by direct commands from a robot. By giving direct speech without any associated negative social outcomes, a robot helper might be an ideal sort of helper. The robot would be able to be clear and direct in its communication without insulting its listeners. This work investigated whether robot helpers could avoid the negative outcomes associated with direct speech, but Experiment 4 offers evidence that robots cannot avoid creating a negative impression when they give direct commands. Despite the practical limitations of this robot as a source of instrumental help, its language was evocative of similar responses as the human helpers. Thus, this thesis articulates the need for robot helpers to address the social and emotional needs of listeners with their help messages. Robot helpers are currently being developed for a wide variety of interactions. Although these interactions are being conceptualized as social interactions, robot designers thus far have avoided the possibility that robot helpers might create anxiety or be insulting. This thesis suggests these negative outcomes are possible for robot helpers and offers specific communication strategies to begin to alleviate these concerns. Fortunately, the expressions of politeness tested in Experiment 4 provide evidence that expressions of politeness in robots are able to create positive impressions when used by robot helpers.

In addition to robot helpers, this finding may be of more general application to a larger class of assistive technology. Increasingly, a variety of assistive devices are using natural language interfaces, and the need for sensitivity in those applications is essential as well. It seems likely that any conversational interface embodied in a robot or software agent should approach their listeners with a basic acknowledgment of the principles of politeness. When designing conversational interfaces for devices where the voice is not embodied, the application of politeness is less clear. It is important for future work to consider how social responses to conversational interfaces might be minimized generally, given that not all social responses to social agents are positive. Simple interactions, such as in-car navigation or talking home appliances, might choose to reduce their vocabulary and de-emphasize their humanness in an attempt to limit the extent to which the conversational technology is perceived to be a social agent, with all the rights and responsibilities therein. To what extent this is possible remains an open research question. While this thesis suggests that politeness strategies are beneficial for embodied agents, such as robots, it does not preclude the possibility that other conversational technologies might be able to avoid the pitfalls of negative responses to help by limiting the social cues displayed by the conversational agent.

Designing for Social Intelligence

This thesis examines two common assumptions in existing human-robot interaction research. First, this thesis questions the implicit assumption that social responses to robots are equivalent to social responses to humans. By observing a human control case, this thesis finds interesting opportunities where responses to humans and robots are not identical. Second, this thesis questions the implicit assumption that modeling a robot's behavior on human behavior will produce positive outcomes. Simply advocating the implementation of social behavior is not sufficient for the design of successful human-robot interaction. The case of help giving is a concrete example of the range of positive and negative outcomes that potentially follow from human behavior. Successful or not, human help-giving behavior is social. The questioning of these assumptions paves the way for a more sophisticated conceptualization of social intelligence, one that considers the complexity of human emotional responses and the range of opportunities in the design of a robot's behavior.

This thesis also allows us to reflect on the current research approach to the design of social intelligence. The Computers as Social Actors theory has been greatly influential in the human-robot interaction community and has been largely used as evidence that robots should mimic human social behavior. The general approach might be categorized as people respond in social ways to robots, therefore robots should behave like people. But this conclusion does not necessarily follow from the premise. Are people the definitive model of social interaction? Are there situations in which people would not be welcome helpers? Considering the baking task again, what if the kitchen was instrumented such that a program could understand when help was needed and display an appropriate video demonstration on a nearby monitor? Not explored in this thesis, but certainly of interest, would be an exploration of different representations of helpers, to investigate when social intelligence is advantageous and when it is a liability.

If the situation requires social interaction, then how can technology developers understand which aspects of social behavior to incorporate into their design? When mimicking human behavior is appropriate, it is still not sufficiently directive. Which humans? Which behaviors? This thesis makes a simple observation that people are different, and it is not always possible to identify a single expert in one's domain and model a robot's behavior on that expert. Another approach to developing a model of a robot's behavior would be to model the "average person," but this approach is problematic as well. As the observations of this thesis describe, people, even the majority of people, are not perfect communicators. Human communicators are satisfying a number of different constraints simultaneously and cannot comprehensively process all the informational and emotional needs of their listeners.

How then can the research community approach the notion of social intelligence? I suggest three opportunities. First, designers of conversational technologies can attempt to limit the extent to which the conversational agent or robot is seen as a fully capable social partner. By limiting vocabulary and by designing the physical form of the robot carefully, designers may be able to lower expectations and avoid some of the challenges involved in having to design the social behaviors perfectly. Second, social intelligence could be designed by using the exact speech of a single individual, with all their nuance and idiosyncrasy. This kind of conversational interface would generate certain negative outcomes, much as any single person might be annoying at times, but listeners would be able to respond to this kind of interface much as they do with the other people in their lives. Finally, modeling human behavior could be done, much like it is done in this thesis, by trying to understand the influence of specific linguistic features and attempting to design an ideal communicator, a conversational agent that attends to a listeners' information needs as well as their emotional needs with appropriateness and sensitivity. This is a costly process of observation and testing, and I do not believe there is any shortcut in using this approach, but the promise of this research is in understanding communication processes to a degree where conversational agents could be more effective and more sensitive than human conversants. It is a long way off, but it is an interesting direction nonetheless.

Future Work

Perhaps the most important future direction for research on robot helpers is to place these robot helpers in the field, where long-term, situated impact might be studied. At present, robot helpers are a technology of the imagined future, and they are generally not sufficiently robust for testing in people's homes, workplaces, or in medical environments. This is a critical limitation of the current work, and only time will reveal how robot helpers will be perceived both socially and emotionally when the novelty has worn off and these helpers are interacting with people in their daily lives.

To expand this work, I believe the next critical step is to understand individual differences between different kinds of help recipients. This work has examined help-giving behavior around a baking task, something that people are generally motivated to do. Our daily activities and interactions, however, are a combination of desires and obligations. Our orientation to our activity is likely to have a profound impact on our reactions to receiving help. So in trying to understand help giving behavior, I believe the crucial next step is conceptualizing people's attachment to their activities. For some people and some activities, help-giving will not be an especially sensitive communicative act. But consider a retired doctor who has been diagnosed with diabetes receiving training about monitoring his glucose level. This is a situation in which a help giver would be careful not to insult the retired doctor's intelligence, while with another listener the help giver might be more directive and do more elaborate explaining. In this situation, it is the help recipients' orientation to the task that is important. If we could understand how a help recipient was approaching the task, how important the task was to the help recipients' identity, we would be better able to provide sensitive help.

Finally, the descriptions of the video clips viewed in Experiment 4 raise some interesting possibilities for future investigation of people's responses to robots. In Experiment 4, I found that people used more abstract language to describe human helpers than robot helpers. This may be an indication that people's preconceptions about robots are influencing their interpretations of the robot's behavior, as articulated by Fischer (2006b). When people see the robot as an abstract social agent, capable of being an "aggressive helper" or a "considerate helper," then the strategies the robot uses will be particularly salient and influential. When people see the robot as doing a task it was designed to do, then it is possible that the robot's politeness strategies will be less influential. This is speculative, of course, and requires further investigation to understand how people's linguistic descriptions of robots interact with the way impressions of robots are formed. In future work, I would like to gather a much larger corpus of people's descriptions of robots to further analyze this connection.

Conclusion

This thesis considers the social and emotional impact of receiving help from a robot. There is a great deal of work that remains to be done in order to understand how we might, one day, live and work alongside robot helpers, but in the meantime this work hopes to convince technologists of the importance of social and emotional outcomes when designing interactions around help giving. This thesis describes the communication strategies of human help givers and draws attention to the rich range of behaviors employed by human help givers used to communicate sensitively with help recipients. Some of these strategies may be appropriate for robots and others may not, but, regardless, this work demonstrates that robot helpers will evoke a range of social reactions with their help messages. Therefore, the communication behavior of robots is a promising place to start when considering the design of an ideal robot helper.

Bibliography

Andersen, G. (2000). The role of the pragmatic marker like in utterance interpretation. In G. Andersen & T. Fretheim (Eds.), *Pragmatic Markers and Propositional Attitude*. Philadelphia: John Benjamins.

Behan, J., & O'Keefe, D. (2008). The development of an autonomous service robot. Implementation: "Lucas" the library assistant robot. *Intelligent Service Robots, 1* (1), 73-89.

Breazeal, C. (2003). Towards Sociable Robots. *Robotics and Autonomous Systems, 42*, 167-175.

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.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York: Cambridge University Press.

Burgoon, J. (1978). A Communication Model of Personal Space Violations: Explication and an Initial Test. *Human Communication Research, 4*(2), 129-142.

Burgoon, J., Le Poire, B. A., & Rosenthal, R. (1995). Effects of Preinteraction Expectancies and Target Communication on Perceiver Reciprocity and Compensation in Dyadic Interaction. *Journal of Experimental Social Psychology, 31*, 287-321.

Caffi, C. (1999). On mitigation. *Journal of Pragmatics, 31*(7), 881-909.

Canary, D. J., & Spitzberg, B. H. (1987). Appropriateness and effectiveness perceptions of conflict strategies. *Human Communication Research, 14*, 93-118.

Clark, H. H. (1996). *Using Language*: Cambridge University Press.

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

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

Coenen, L. H., Hedebouw, L., & Semin, G. R. (2006). *Measuring Language Abstraction: The Linguistic Category Model Manual*.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by presence of others. *Journal of Personality and Social Psychology, 9*, 245-250.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society, 362*(1480), 679-704.

DePaulo, B., & Fisher, J. (1980). The costs of asking for help. *Basic and Applied Social Psychology, 1*(1).

Drurian, P. R., & DePaulo, B. (1977). Asking a child for help. *Social Behavior & Personality, 5*, 33-39.

Fischer, K. (2006a). *Approaches to Discourse Particles*: Emerald Group Publishing.

Fischer, K. (2006b). *The Role of Users' Preconceptions in Talking to Computers and Robots*. Workshop on How People Talk to Computers, Robots, and other Artificial Communication Partners.

Fisher, J., Nadler, A., & Whitcher-Alagna, S. (1982). Recipient reactions to aid. *Psychological Bulletin, 91*(1), 27-54.

Flynn, F. J. (2006). How much is it worth to you? Subjective evaluations of help in organizations. *Research in Organizational Behavior, 27*, 133-174.

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: Effects of computers that flatter. *International Journal of Human-Computer Studies, 46*(5), 551-561.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3-4), 143-166.

Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics, 34*, 727-747.

Fraser, B. (1980). Conversational mitigation. *Journal of Pragmatics, 4*(4), 341-350.

Freier, N. G. (2008). *Children attribute moral standing to a personified agent.* Proceedings of the ACM Conference on Human Factors in Computing, CHI 2008.

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

Galinsky, A., Ku, G., & Wang, C. (2005). Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations, 8*(2), 109-124.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004a). *Action as language in a shared visual space.* Proceedings of the Conference on Computer-Supported Cooperative Work, CSCW 2004, Chicago, IL.

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

Gharpure, C., & Kulyukin, V. (2008). Robot-assisted shopping for the blind: issues in spatial cognition and product selection. *Intelligent Service Robots, 1*(3).

Giles, H., Coupland, N., & Coupland, J. (1991). Communication, context, and consequence. In H. Giles, J. Coupland & N. Coupland (Eds.), *Contexts of* *accommodation: developments in applied sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

Gockley, R., Forlizzi, J., & Simmons, R. (2006). *Interactions with a Moody Robot*. Proceedings of the International Conference on Human Robot Interaction, HRI 2006.

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

Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Doubleday.

Greenberg, M., & Westcott, D. (1983). Indebtedness as a mediator of reactions to aid. In J. D. Fisher, A. Nadler & B. DePaulo (Eds.), *New directions in helping: Vol. 1 Receipient reactions to aid* (pp. 51-84).

Hayaski, K., Sakamoto, D., Kanda, T., Shiomi, M., Koizumi, S., Ishiguro, H., et al. (2007). *Humanoid robots as a passive-social medium: a field experiment at a train station*. Proceedings of the International Conference on Human Robot Interaction, HRI 2007.

Holmes, J. (1986). Functions of you know in women's and men's speech. *Language in Society, 15*(1), 1-22.

Holtgraves, T. (2002). Face management and politeness. In T. Holtgraves (Ed.), *Language as Social Action: Social Psychology and Language* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum.

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

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

Jucker, A. H., & Smith, S. W. (1998). And people just you know like wow: Discourse markers as negotiating strategies. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers: Descriptions and Theory* (pp. 171-201). Philadelphia: John Benjamins.

Kahn Jr., P. H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., et al. (2007). What is human?: Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies, 8*(3), 363-390.

Kang, K. I., & Mataric, M. (2005). *A hands-off physical therapy assistance robot for cardiac patients*. Proceedings of the IEEE International Conference on Rehabilitation Robotics, ICORR 2005.

Kidd, C. D., & Breazeal, C. (2007). *A robotic weight loss coach*. Proceedings of the Twenty-Second Conference on Artificial Intelligence, AAAI 2007.

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., 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.

Lee, E. J., Nass, C., & Brave, S. (2000). *Can computer-generated speech have gender?: an experimental test of gender stereotype*. Extended Abstracts on Human Factors in Computing Systems, CHI 2000.

Lee, F. (1997). When the Going Gets Tough, Do the Tough Ask for Help? Help Seeking and Power Motivation in Organizations. *Organizational Behavior and Human Decision Processes, 72*(3), 336-363.

Lee, F. (2002). The social costs of seeking help. *Journal of Applied Behavioral Science, 38*(1).

Lee, S., Kiesler, S., Lau, I. Y., & Chiu, C. (2005). *Human mental models of humanoid robots.* Proceedings of the IEEE International Conference on Robotics and Automation, ICRA 2005, Barcelona, Spain.

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

MacDorman, K. F., & Cowley, S. J. (2006). *Long-term relationships as a benchmark for robot personhood*. Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication.

McCrosky, J. C. (1966). Speech Monographs. *Scales for the measurement of ethos, 33*, 65-72.

Morkes, J., Kernal, H. K., & Nass, C. (1998). *Humor in task-oriented computer-mediated communication and human-computer interaction*. Proceedings of the Conference on Human Factors in Computing Systems, CHI 1998.

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009). *Footing in Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues.*

Proceedings of the International Conference on Human Robot Interaction, HRI 2009.

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). *Nonverbal Leakage in Robots: Communication of Intentions through Seemingly Unintentional Behavior*. Proceedings of the International Conference on Human Robot Interaction, HRI 2009.

Nadler, A. (1980). Good looks do help: Effects of helper's physical attractiveness and expectations for future interaction on help-seeking behavior. *Personality and Social Psychology Bulletin, 6*, 378-384.

Nadler, A. (1990). Help-seeking behavior as a coping resource. In M. Rosenbaum (Ed.), *Learned resourcefulness: On coping skills, self-control, and adaptive behavior* (pp. 127-162): Springer.

Nadler, A., Peri, N., & Chemerinski, A. (1985). Effects of opportunity to reciprocate and self-esteem on help-seeking behavior. *Journal of Personality, 53*(1), 23-35.

Nakano, Y., Reinstein, G., Stocky, T., & Cassell, J. (2003). *Toward a model of face-to-face grounding.* Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? A test of the similarity-attraction hypothesis. *Journal of Experimental Psychology: Applied, 7*, 171-181.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2002). Towards robotic assistants in nursing homes: challenges and results. *Workshop on Robot as Partner: An Exploration of Social Robots, IEEE International Conference on Robots and Systems*.

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 International

Workshop on Robots and Human Interactive Communication, RO-MAN 2005, Nashville, TN.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Sacks, H., Schegloff, E., & Jefferson, G. (1978). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2009). *How to Approach Humans? Strategies for Social Robots to Initiate Interactions*. Proceedings of the International Conference on Human Robot Interaction, HRI 2009.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

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

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

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

Schourup, L. (1999). Discourse markers: tutorial overview. *Lingua, 107*, 227-265.

Siegel, M. E. A. (2002). Like: The discourse particle and semantics. *Journal of Semantics, 19*, 35-71.

Smith, J. D., Vertegaal, R., & Sohn, C. (2005). *ViewPointer: Lightweight calibration-free eye tracking for ubiquitous handsfree deixis.* Proceedings of the Conference on User Interface Software & Technology, UIST 2005, Seattle, WA.

Staudte, M., & Crocker, M. W. (2009). *Visual Attention in Spoken Human-Robot Interaction*. Proceedings of the International Conference on Human Robot Interaction, HRI 2009.

Stubbe, M., & Holmes, J. (1995). You know, eh and other exasperating expressions: an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication, 15*(1), 63-88.

Tapus, A., Tapus, C., & Mataric, M. (2008). User-Robot Personality Matching and Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy. *Intelligent Service Robotics Journal, 1*(2), 169-183.

Walters, M. L., Dautenhahn, K., Woods, S. N., & Koay, K. L. (2007). *Robotic etiquette: Results from user studies involving a fetch and carry task.* Proceedings of the International Conference on Human Robot Interaction, HRI 2007.

Warner, R. M., & Sugarman, D. B. (1986). Attributions of personality based on physical appearance, speech, and handwriting. *Journal of Personality and Social Psychology*, 792-799.

Weimann, J. M. (1977). Explication and test of a model of communicative competence. *Human Communication Research, 3*, 195-213.

Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). *Developing a Model of Robot Behavior to Identify and Appropriately Respond to Implicit Attention-Shifting.* Proceedings of the International Conference on Human Robot Interaction, HRI 2009.

Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., & Kuzuoka, H. (2008). *Precision timing in human-robot interaction: coordination of head movement and utterance.* Proceedings of the Conference on Computer Supported Cooperative Work, CSCW 2008.

Yamazaki, K., Kawashima, M., Kuno, Y., Akiya, N., Burdelski, M., Yamazaki, A., et al. (2007). *Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings.* Proceedings of the European Conference on Computer Supported Cooperative Work, ECSCW 2007.

Appendix A

Each numbered column references discreet helping exchanges (a series of turns surrounding the exchange of a single piece of information). Discourse markers spoken by helper are listed above the numbered line; hedges spoken by helper are listed below the numbered line.

Session A

I mean yeah	so			oh	like	so			just like	just like	just like so	so		actually oh
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
I figure							I guess							

actually oh		just	just	like so yeah			like yeah	actually just so yeah			so well	basically just like so yeah		so yeah	actually oh
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	probably						almost I guess I think			might			I think		

Session B

well		yeah	yeah	so	oh	actually just so		like so	so			so	so yeah		well						so	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
																	might	probably				

Session C

												just			
												like			
				so								yeah			
basically			oh	well	like	well	just	yeah		oh		you know			yeah
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
					might										

																		just
	like		just	well			so			just							just	you know
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35

Session D

																well			
																oh		yeah	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
																I think			I think
																			might

																just	
oh	just	yeah			oh			yeah		actually		so	yeah	yeah	yeah	yeah	
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38

Session E

basically just like	actually	just so well yeah oh	like	actually	just	like yeah	just like	so	like	just like	just like so oh	oh		just yeah	just	just like	actually so yeah
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
probably		I think		might probably	maybe		maybe		maybe		I think	might			might probably		I guess maybe

like so		actually just like well	just like yeah	so yeah	so	just like so	just	just	actually just so	just like	so	just so	just yeah you know		I mean just	
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
I guess	maybe		might		probably									probably		I think maybe

Session F

																			actually
								just											just
								like											like
		like						so		like		actually							so
yeah	like	so		yeah		actually		yeah	like	yeah		just	like		well			just	oh
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		I think				might						probably	probably				I think		
		maybe															probably		

												actually				actually		
												I mean				I mean		
												like		actually		like		
		just	actually	just			actually			just	actually	so		just		well		
oh	like	like	like	oh		like	well	like	yeah	so	like	yeah		like		oh	so	yeah
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
	probably		probably		probably		I think		might	probably		probably		maybe	probably	I guess		I guess
																probably		maybe