Designing Embodied Cues for Dialogue with Robots

Bilge Mutlu

■ *Of all computational systems, robots are unique in their ability to afford embodied interaction using the wider range of human communicative cues. Research on human communication provides strong evidence that embodied cues, when used effectively, elicit social, cognitive, and task outcomes such as improved learning, rapport, motivation, persuasion, and collaborative task performance. While this connection between embodied cues and key outcomes provides a unique opportunity for design, taking advantage of it requires a deeper understanding of how robots might use these cues effectively and the limitations in the extent to which they might achieve such outcomes through embodied interaction. This article aims to underline this opportunity by providing an overview of key embodied cues and outcomes in human communication and describing a research program that explores how robots might generate high-level social, cognitive, and task outcomes such as learning, rapport, and persuasion using embodied cues such as verbal, vocal, and nonverbal cues.*

One of the most promising concepts that emerged in the last two decades of computer science research is *embodied interaction* — representing computational intelligence in and offering interaction through physical or simulated ''bodies.'' Such representations vary from physical artifacts (such as tangible interfaces) to biological forms (such as humanlike agents and robots) and offer templates for understanding and interacting with complex computational systems (Ullmer and Ishii 2000, Cassell 2001, Breazeal 2003). Physically embodied humanlike attributes provide particularly rich representations with which people are intimately familiar, and systems that draw on these representations promise significant social and task outcomes in key application domains from education (Kanda et al. 2004) to collaboration (Bluethmann et al. 2003) to rehabilitation (Dautenhahn and Werry 2004).

Embodied interaction offers not only a familiar template for making sense of and interacting with computational systems but also an opportunity to achieve the most effective forms of human interaction for achieving positive social and task outcomes in such domains. How do *the most effective* teachers improve student learning? How do *the most effective* speakers capture their audiences? How do *the most effective* personal coaches motivate people to achieve their goals? Research in human communication has shown that brief observations of the bodily, facial, speech, and vocal cues that people display predict these outcomes (Ambady and Rosenthal 1992). Seemingly subtle differences in such cues of interaction partners shape outcomes such as perceptions of their attitudes (Mehrabian 1967), the persuasiveness of their messages (Segrin 1993), and their performance in collaborative work (Burgoon et al. 2002).

This tight coupling between communicative cues and key social, cognitive, and task outcomes opens up a space for designing effective interactions for robots to elicit particular positive

 ISSN 0738-4602

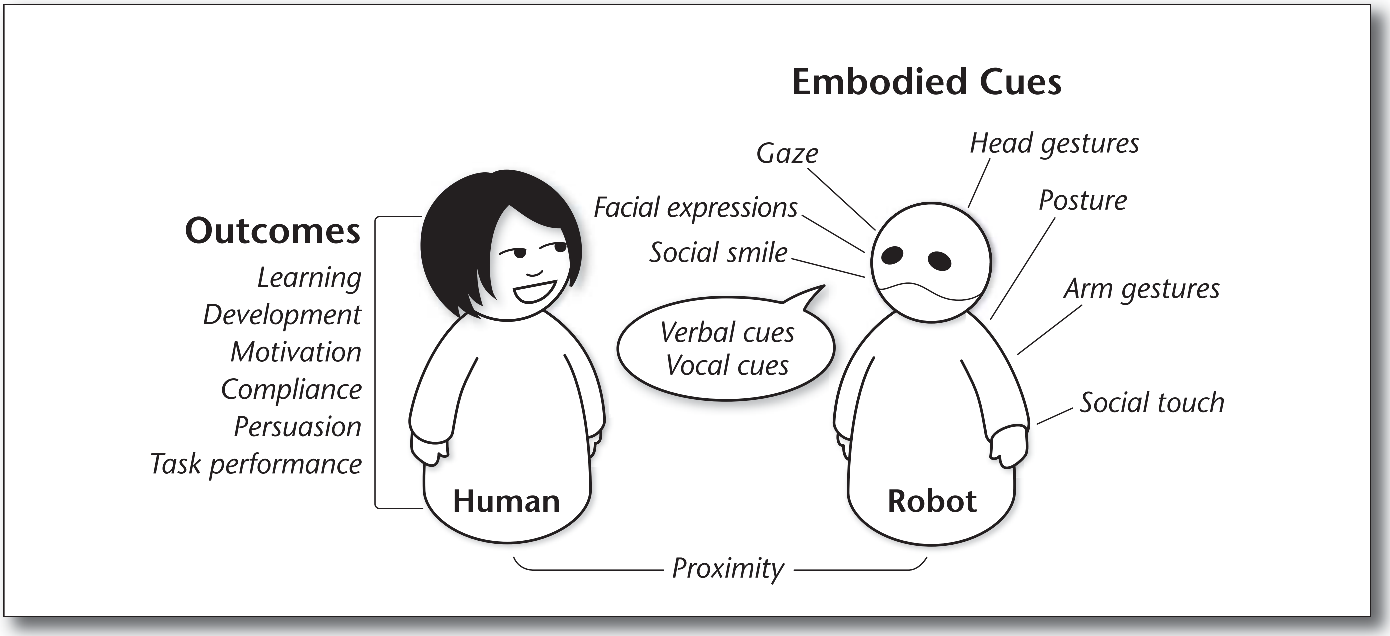

Figure 1. Key Cues and Outcomes in Embodied Human-Robot Interaction.

outcomes in key domains of human life. The exploration of this design space must necessarily include careful studies that iteratively manipulate how embodied cues should be used and combined in order to achieve the most effective interactions. These studies must also seek to understand the limitations in the extent to which robots might achieve outcomes that human interactions generate in human-robot interaction. This understanding will enable designers to create robots that serve as effective teachers, motivators, and collaborators. Which embodied cues are important? What are key outcomes? How well do they work when robots use them? This article seeks to answer these question by providing an overview of key embodied cues and outcomes and outlining a research program that aims to understand how robots might use embodied cues to achieve significant outcomes in human-robot interaction.

Which Embodied Cues?

Human communication is enabled by a number of embodied cues, which include behaviors such as gestures, vocal stress, and linguistic markers interlocutors produce — with or without intent — that form or shape the content of their messages. Even seemingly disembodied speech and vocal cues originate in the human body and co-occur with other embodied cues. While each cue may seem subtle and insignificant in isolation, the cues occur and co-occur in complex patterns to make up a system of communication and elicit significant social and task outcomes. This section provides brief overviews of some of the embodied communicative cues that human communication researchers have considered over the last several decades and that serve as key design variables for effective dialogue with robots (figure 1). It focuses on cues that appear in situated dialogue in dyads or small groups and that shape the flow of communication, the participants' perceptions of others, and individual or shared task outcomes.

Verbal Cues. Spoken language involves an immensely rich set of embodied communication cues that shape social and task outcomes. These cues enable coordinated actions toward achieving common ground between communication partners (Clark 1996). The level of coordination in these activities shape joint task outcomes (Hancock and Dunham 2001). These cues also serve as indirect acts that shape the social interaction between the participants (Goffman 1969, Brown and Levinson 1988). For instance, by changing the semantic structure of their speech, or including specific markers such as "please," speakers can increase how polite others perceive them (Brown and Levinson 1988). Similarly, speakers who use definite references such as "the Anderson Marketplace" as opposed to "a marketplace" might be perceived as having more expertise on the topic (Clark and Marshall 2002). This perception of expertise may affect the persuasive ability of these speakers, as research shows that personal credibility significantly shapes persuasion (Hartelius 2008).

Vocal Cues. Another set of embodied communication cues connected with verbal cues includes vocal, paralinguistic, or prosodic cues such as vocal pitch, rhythm, volume, tempo, voice quality, and the distribution and length of pauses (Wennerstrom 2001). These cues facilitate processing and understanding of spoken words, syntactic structure, discourse structure (Cutler, Dahan, and Van Donselaar 1997) and communicate affect, attitudes, and intentions (Scherer, Ladd, and Silverman 1984). Using vocal cues, speakers can achieve a wide range of functions from facilitating turn-taking (Duncan 1972) to polite speech (Ambady et al. 1996).

Gaze. Gaze cues are composed of the orientations of the eyes, head, and body and communicate the direction of attention (Frischen, Bayliss, and Tipper 2007). These cues appear in context-specific patterns that facilitate a number of communicative processes including mutual and averted gaze, gaze following, joint and shared attention, and the theory of mind (Emery 2000).

Gestures. Gestures such as arm, hand, and head movements form another set of salient embodied communication cues (Krauss 1998, McClave 2000). Arm gestures appear in a number of forms — particularly, iconic, metaphoric, beat, cohesive, and deictic gestures — and embody specific symbolic meanings (McNeill 2005). Speakers use iconic and metaphoric gestures to communicate concrete objects and abstract ideas, respectively, through pictorial representations. Deictic gestures involve pointing toward information in the environment. Beat gestures indicate that the verbal cues that they accompany are significant. Cohesive gestures connect parts of speech that are thematically related across speech segments.

Head gestures are displayed by speakers and listeners alike, primarily in the form of nods and shakes (Hadar et al. 1984;, McClave 2000). When produced by listeners, nods and shakes might signal agreement and disagreement, respectively (Poggi, D'Errico, and Vincze 2010), while also serving as back-channel requests, deictic references, or expressions of uncertainty (McClave 2000).

Proximity. The physical distance and orientation people maintain between themselves and others serve as a salient embodied cue (Hall 1966, Kendon 1973). Orientation and distance cues also interact with each other; people maintain different amounts of distance between themselves and others based on their relative orientations (Hayduk 1981).

Facial Expressions. Expressions of emotional and other internal states through movements of the face form another set of prominent embodied cues. While the relationship between these expressions and internal states is not straightforward, researchers agree that a person's facial expressions shape people's perceptions of the person (Ekman 1993; Russell, Bachorowski, and Fernández-Dols 2003). Some of these expressions are given as communicative cues, while some are given off as byproducts of other behaviors (Goffman 1959).

Social Smile. In conversations, participants frequently display a particular type of facial expression: social smiles. Smiling in this context expresses understanding and agreement, serving as a back-channel and improving conversational efficiency (Brunner 1979). Individuals who smile more are considered to be more socially competent (Argyle 1988, Otta et al. 1994).

Social Touch. Social touch involves physical interaction between communication partners such as hand-holding, handshake, and touching the forearm, shoulder, or face and serves as a uniquely embodied and potent communicative cue (Jones and Yarbrough 1985, Gallace and Spence 2010). The various forms of touch express dominance as well as intimacy, immediacy, affection, and trust (Mehrabian 1972, Montagu and Matson 1979, Burgoon 1991).

Posture. Posture cues such as openness of arms and legs, arms-akimbo position, and trunk relaxation serve as embodied cues, providing information on the attitudes of and status relationships between communication partners (Mehrabian 1969) and shaping how they perceive each other (Osborn 1996).

The set of cues described here forms a rich design space for creating embodied cues for robots. The next section illustrates how the effective use of these cues might generate significant social, cognitive, and task outcomes in different domains of human communication.

What Key Outcomes?

Embodied communication cues, such as those described earlier, activate key social and cognitive processes and, in turn, elicit positive outcomes. This indirect relationship between cues and outcomes poses an opportunity for designing behavioral mechanisms for robots that seek to achieve social, cognitive, and task outcomes. This section provides an illustrative set of key domains of social interaction in which embodied communication cues are known to yield such significant outcomes. These domains involve social situations in which a communication partner seeks to improve individual or shared outcomes in social (for example rapport), cognitive (for example learning), and task (for example task performance) measures. These outcomes inevitably pose some overlap as embodied cues activate social and cognitive processes that are shared across domains. They also have limited generalizibility across social situations, as factors such as cultural context significantly affect the production and perception of embodied cues and, therefore, their outcomes.

Learning and Development. The key role of embodied cues has been particularly emphasized in the context of learning and development. Education researchers have explored the role of teachers' embodied cues in achieving higher levels of immediacy — the degree of physical and psychological closeness between people (Mehrabian 1966) — and teaching effectiveness. These cues include gaze, gestures, proximity, posture, facial expressions, touching, and vocal tone and expressions (Richmond, Gorham, and McCroskey 1987). Research to date shows a strong, positive relation between perceptions of teachers' nonverbal immediacy and students' evaluations of the teacher, the class, and self-reported learning (Harris and Rosenthal 2005). Specific manipulations in these cues lead to significant learning outcomes. For example, classroom research has shown increased teacher gaze toward students to improve information recall across different student populations from primary school to college (Otteson and Otteson 1980, Sherwood 1987). Similarly, teacher gestures are found to improve student learning by grounding verbal concepts in physical references (Graham and Heywood 1975, Thompson and Massaro 1994, Alibali and Nathan 2007).

Embodied communication cues also activate key developmental processes. Research on language acquisition has found that infants showed significantly more learning when they were exposed to a new language through social interaction than when they received the same amount of exposure through video or audio (Kuhl 2007). Research in this area also emphasizes the role of specific cues in language learning. For instance, pairing a verbal reference to an object with a gaze cue toward the object might facilitate the acquisition of the name of that object (Baldwin 1995). In fact, developmental research has found that infants' ability to follow gaze cues at six months correlates with vocabulary size at 18 months (Morales et al. 2000).

Attention and Engagement. Embodied cues of a partner can also evoke attention and increase engagement in people. Classroom research suggests that teachers consciously and explicitly use gaze cues to attract the attention of their students (Woolfolk and Brooks 1985). Eye contact is found to be one of the main factors to increase the efficacy of verbal reprimands in the classroom (Van Houten et al. 1982). Teachers' use of embodied cues might also indicate their level of experience and effectiveness; experienced and/or effective teachers and inexperienced and/or ineffective teachers show significant differences in the frequency of direct eye contact with students during the first week of class (Brooks 1985).

Embodied cues can also indicate attraction toward people or objects. Research on gaze cueing has shown that participants find partners who establish mutual gaze with them more likable and attractive than those who avert their gaze (Mason, Tatkow, and Macrae 2005). Similarly, participants evaluate objects that their partners are looking toward more favorably than they evaluate those that their partners are looking away from (Bayliss et al. 2006).

Attention evoked by embodied cues can also affect economic decisions (Haley and Fessler 2005; Bateson, Nettle, and Roberts 2006; Burnham and Hare 2007). Research in this area has shown that individuals pay nearly three times more for drinks when a pair of eyes were placed on a donation-based vending machine then when images of flowers were placed on the machine (Bateson, Nettle, and Roberts 2006). Similarly, people contribute more in public-goods games when they see images of a robot on their computer screen (Burnham and Hare 2007) or when a pair of schematic eyes are placed in their computer background (Haley and Fessler 2005).

Research has also shown that increased attention through embodied cues can improve task outcomes. For instance, in a digit-encoding task, participants performed better when the instructors made as much eye contact as possible while reading the instructions than they did when they made as little eye contact as possible (Fry and Smith 1975).

Motivation, Compliance, and Persuasion. Research on the social and task outcomes of embodied communication cues has also explored their effects on motivation, compliance, and persuasion. A metareview of 49 studies that explored the effects of embodied cues including gaze, touch, and proxemity has found that these cues consistently increased compliance with the requests or messages of a partner (Segrin 1993). In the context of public speaking, vocal and nonverbal cues including facial expressions, gaze, gestures, posture, and proximity are strongly associated with the credibility and the persuasiveness of the speaker (Burgoon, Birk, and Pfau 1990).

A large number of studies in health-care, education, and public contexts show that physical touch improves measures of compliance, motivation, and persuasion. In health care, touching by the health-care provider increases patients' compliance with nutritional instructions (Eaton, Mitchell-Bonair, and Friedmann 1986) and disclosure in therapy (Jourard and Friedman 1970). In the classroom, touching on the forearm by a teacher increases the likelihood that students will volunteer for a public demonstration (Guéguen 2004). Finally, research in public contexts has shown that waiters or waitresses increase their gratuity by touching customers (Crusco and Wetzel 1984); interviewers increase response rate to street surveys through touch (Hornik and Ellis 1988); and shopkeepers can solicit interest in their products by touching their customers (Hornik 1992).

Collaboration. Embodied cues facilitate collaborative processes such as grounding, shared attention, and perspective taking and improve joint task outcomes. For instance, information on a partners' gaze direction increases task outcomes in collaborative visual search (Brennan et al. 2008), referential grounding (Liu, Kay, and Chai 2011), and story comprehension (Richardson and Dale 2005). Research on collaborative work has also shown that gestures play a key role in conversational grounding and collaborative task outcomes (Fussell et al. 2004; Kirk, Rodden, and Fraser 2007).

These outcomes highlight the importance of embodied cues in human interaction and underline the potential that robots hold for significant impact through embodied interaction. The next section describes a research program that seeks to explore this potential by studying how robots might use particular embodied cues to generate particular positive outcomes in human-robot interaction.

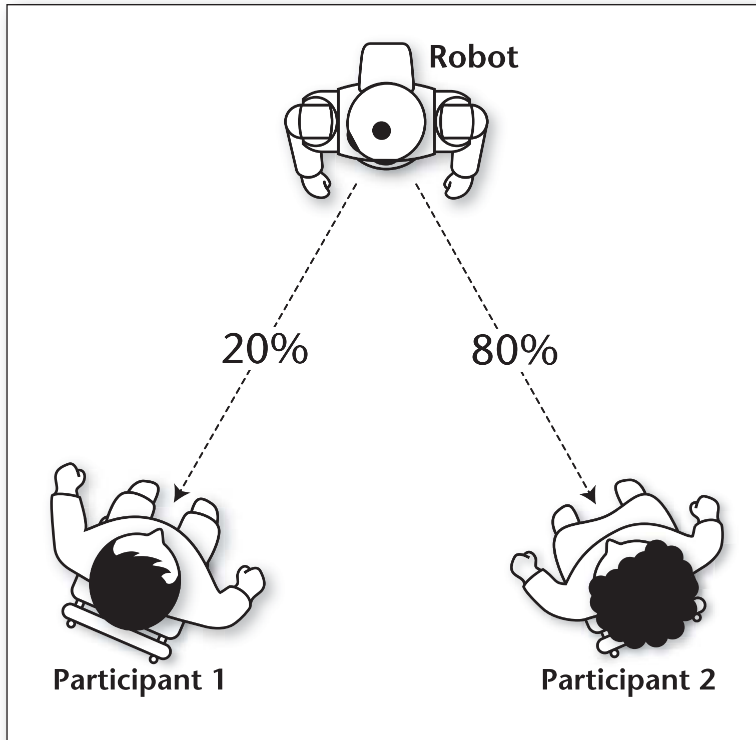

Figure 2. The Setup of the First Study.

Designing Embodied Cues for Robots

The relationship between embodied cues and social, cognitive, and task outcomes promises robot designers opportunities for designing embodied cues for robots to achieve similar outcomes in human-robot dialogue. This section provides brief descriptions of an illustrative set of recent studies from a research program on designing cues for embodied dialogue with robots. The set of studies explores how robot cues such as gaze, proximity, gestures, and verbal and vocal cues might achieve high-level outcomes including improved attention, learning, rapport, compliance, and persuasion. All studies ground the design of the embodied cues of the robot in careful observations of human behavior or findings from existing research on human communication. Because this article focuses on the connection between low-level cues and high-level outcomes, details of the modeling of human behaviors and the interaction design of robot behaviors have been omitted. The reader can find these details in the respective current or future publications on each study.

Gaze, Information Recall, and Liking

The first study explored how a robot's gaze cues might be manipulated to achieve learning outcomes in participants. In the study, two participants listened to a story told by Honda's ASIMO humanlike robot. The robot's gaze behavior followed a partly stochastic, partly rule-based, data-driven model of gaze developed to achieve humanlike gaze shifts that accompany speech (Mutlu, Forlizzi, and Hodgins 2006). The study tested the hypotheses that increased gaze would (1) increase the recall of the details of the story by the participants and (2) improve the participants' overall evaluation of the robot.

The testing of these hypotheses involved manipulating the robot's gaze behavior between participants by changing the total frequency of gaze shifts directed to the two participants to be 80 percent and 20 percent instead of 50 percent for both (figure 2). The independent variables included the frequency of the robot's gaze and participant gender. The dependent variables involved short-term, retained information recall, the participants' affective states, their evaluations of the robot's social and intellectual characteristics, and their evaluations of the task.

A total of 20 college students, 12 males and 8 females, participated in the study in a total of 10 sessions. In each session, one of the participants received 80 percent of the attention of the robot and the other received 20 percent. All participants were native English speakers and their ages ranged from 19 to 33. In the study, following informed consent and a pre-experiment questionnaire, the participants listened to the robot tell a traditional Japanese fable titled "Shita-kiri Suzume" (Tongue-Cut Sparrow) in English. The participants then listened to a second story, Hans Christian Anderson's "The Flying Trunk," on tape as a distractor task. Following the second story, the participants took a

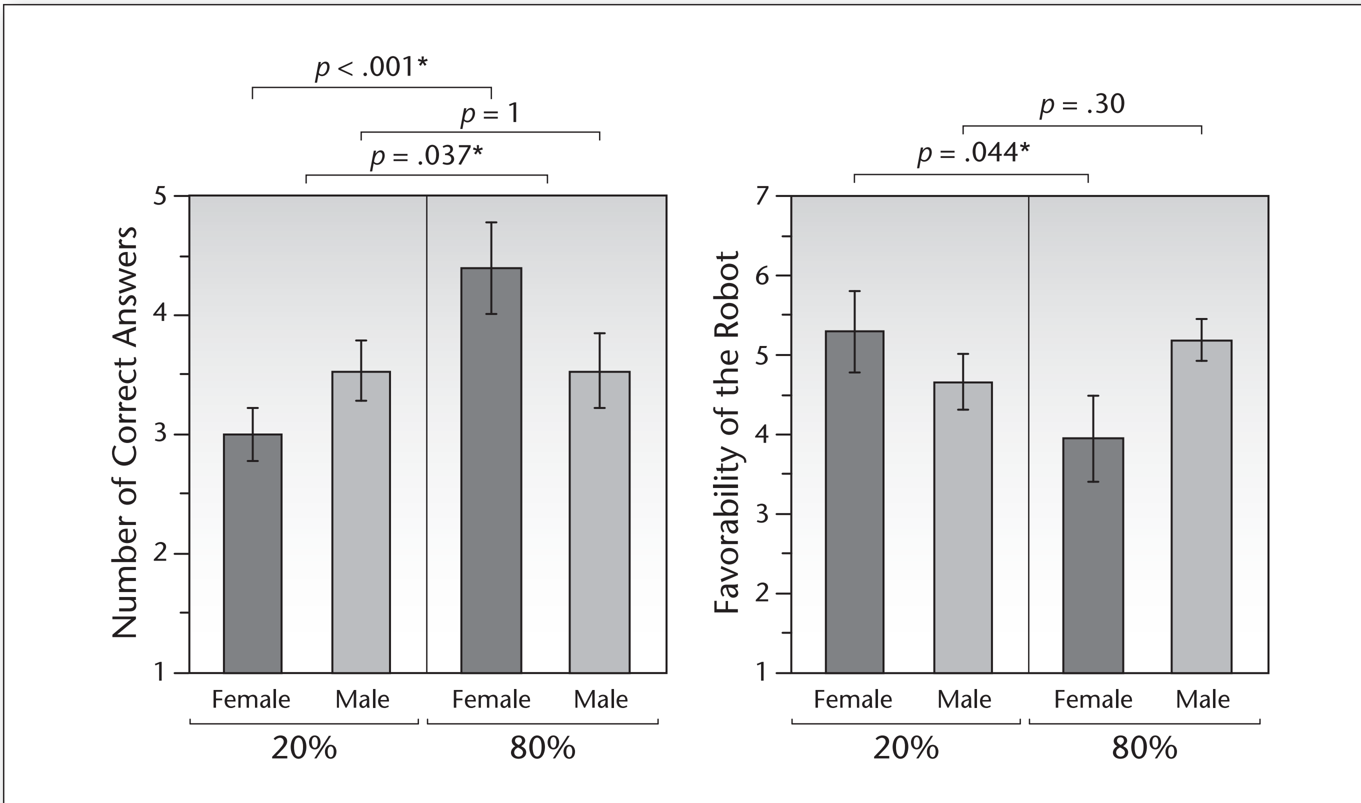

Figure 3. The Number of Correct Answers and the Robot's Favorability Across Gaze Conditions and Genders.

test on the robot's story and answered a postexperiment questionnaire.

Manipulation checks showed that the participants toward whom the robot looked more indicated that the robot looked at them marginally more than those toward whom the robot looked less, $F(1, 16) = 4.29$, $p = .053$. Consistent with the first hypothesis, the results showed that participants who received 80 percent of the robot's gaze had higher information recall than those who received 20 percent of the robot's gaze did, $F(1, 16) = 5.15$, $p = .037$. However, the results also showed a significant interaction between the gaze manipulation and gender, $F(1, 16) = 5.15$, $p = .037$. The effect of the gaze manipulation was significant only for women, $F(1, 16) = 8.58$, $p < .001$, and not for men, $F(1, 16) = 0$, $p = 1$. The results showed no significant effect of gaze or gender on the participants' evaluations of the robot, but indicated a significant interaction, $F(1, 16) = 5.62$, $p = .031$. Females toward whom the robot looked more evaluated the robot less positively than those toward whom the robot looked less did, $F(1, 16) = 4.80$, $p = .044$, while males' evaluations did not differ across conditions, $F(1, 16) = 1.14$, $p = .30$ (figure 3).

The results of the study confirmed the relationship between gaze cues and learning outcomes and showed strong gender effects on this relationship. Increased robot gaze improved information recall in females, but not in males. Contrary to the prediction, increased gaze decreased the favorability of females' evaluations of the robot, but did not affect those of males.

Verbal Cues and Persuasion

The second study investigated how a robot might improve the persuasiveness of its messages using verbal cues, particularly linguistic markers of expertise. To achieve more expertlike speech and, in turn, improve its persuasiveness, Mitsubishi's Wakamaru robot used proper nouns and definite references instead of indefinite references (for example "The Eleanor Art Museum" instead of "the art museum") and references to past experience as opposed to references to information visible in a picture. The study tested two hypotheses: (1) the participants would express a stronger preference toward options presented using expert language than they do toward options presented using nonexpert language and (2) the persuasion outcome would be stronger for women than it would be for men.

In the study, the participants constructed a walking map in a fictional city and listened to information from Wakamaru on the alternative landmarks that they could visit (figure 4). At every intersection, the robot provided information on two similar landmarks (for example two amusement parks), using linguistic markers of expert speech with one and those of nonexpert speech with the other. The independent variables included the presence of expert language markers, manipulated within participants, and participant gender, measured between participants. The main dependent variable was whether participants included the landmarks described with expert language in their walking plans.

Twenty-six college students, 16 males and 10 females whose ages ranged from 19 to 28, participated in the study. All participants were native English speakers. Following informed consent, the experimenter told the participant that the robot was being trained to be a tour guide and that the participant's assistance was needed to evaluate the robot's effectiveness. The participant sat in front of a computer and the robot stood next to the participant and the computer. During the task, the robot sought to establish and maintain common ground with the participant and follow conversational conventions. When the participant reviewed the landmark options and carried out the other experimental tasks, the robot looked toward the participant's screen. When it spoke, the robot maintained eye contact with the participant. After completing the task, the participants provided answers to a postexperiment questionnaire.

The results showed that, consistent with the first hypothesis, the participants included significantly more landmarks described using expert language than those described using nonexpert language in their walking maps, $F(1, 48) = 71.23$, $p < .001$ (figure 5). The analysis also indicated a marginal interaction between the language manipulation and participant gender, $F(1, 48) = 3.79$, $p = .057$. Post-hoc tests showed that expert language affected both women, $F(1, 48) = 43.84$, $p < .001$, and men, $F(1, 48) = 27.40$, $p < .001$, but the effect size was larger for women ($\eta_p^2 = .477$ versus $\eta_p^2 = .363$).

The results of the study supported the prediction; verbal cues of expertise increased the persuasiveness of the information that the robot provided and affected the participants' preferences. These cues affected men and women alike with an overall stronger effect on women. This study demonstrates the effectiveness of verbal cues of expertise in crafting persuasive messages and shows the potential benefit of developing a model of expert speech for robots.

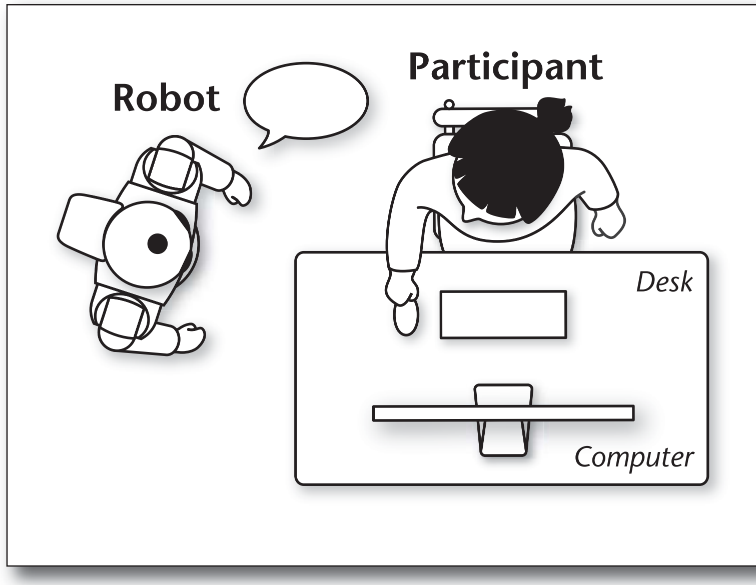

Figure 4. The Setup of the Second Study.

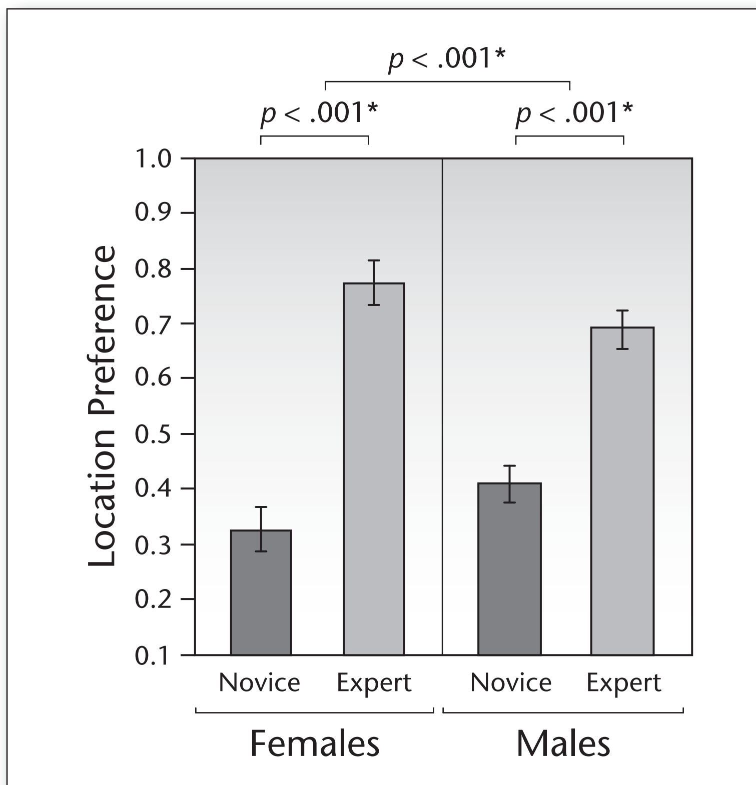

Figure 5. The Probability of Participants Including a Landmark in Their Walking Maps Across Conditions and Genders.

Gaze, Verbal Cues, Rapport, and Distancing

The third study explored how robots might use verbal and nonverbal cues to improve their proxemic relationship with people, affecting outcomes such as distancing, rapport, and disclosure. In this study, Wakamaru varied its gaze behavior and ver-

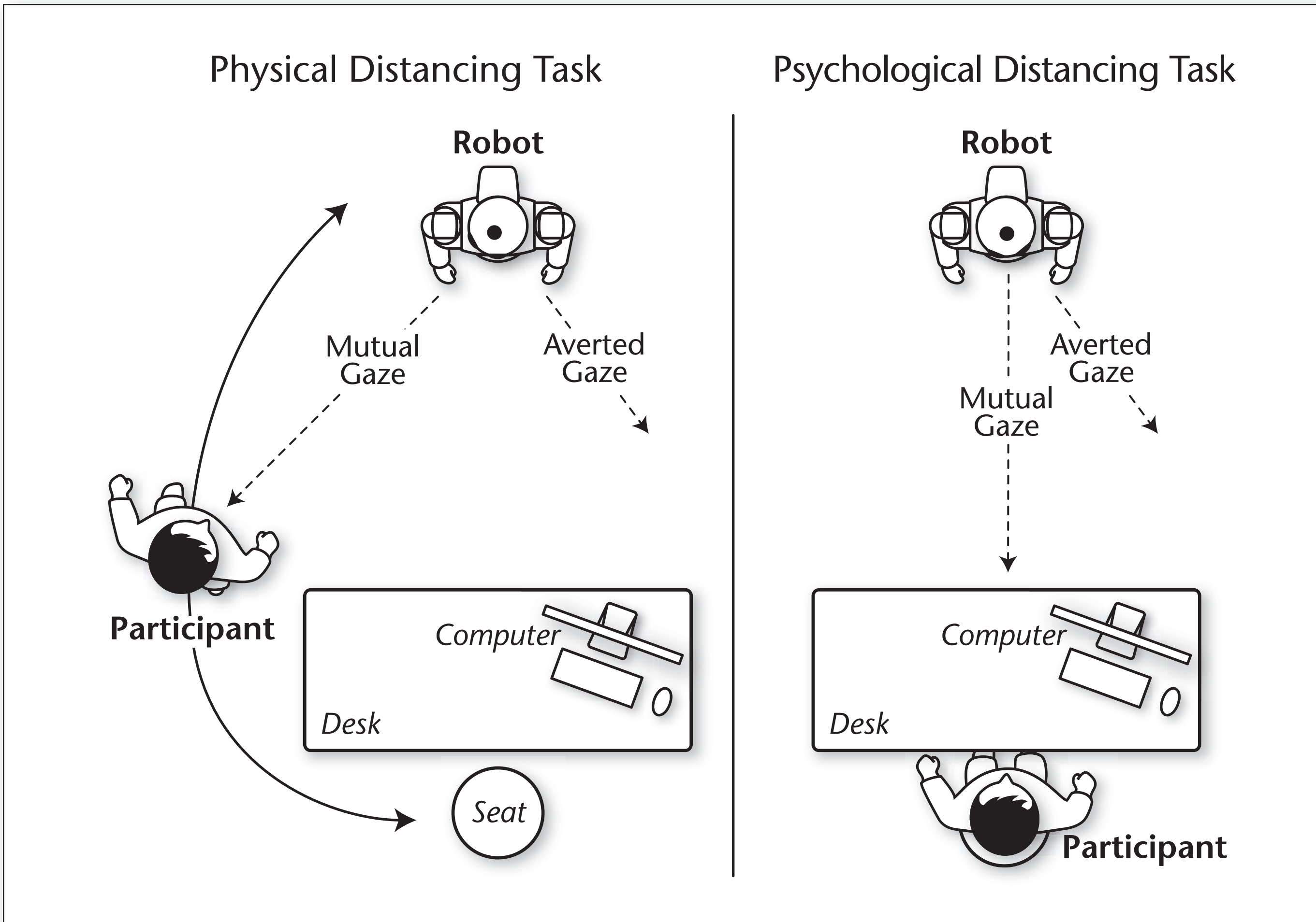

Figure 6. The Setup of the Third Study.

bal politeness cues to shape the participants' physical (for example proximity) and psychological (for example rapport) distancing from the robot, respectively. The study tested three hypotheses: (1) the participants would maintain a greater physical distance with the robot when the robot follows them with its gaze than they do when the robot averts its gaze; (2) the participants with whom the robot does not have rapport would conform to the prediction of the first hypothesis and those with whom the robot has rapport will show the opposite effect; and (3) the participants would disclose less information to the robot when the robot follows them with its gaze than they do when the robot averts its gaze.

In the study, the participants were asked to perform two tasks that required them to control the amount of physical and psychological distancing, respectively (figure 6). The independent variables included the amount of gaze that the robot directed toward the participants (gaze following versus aversion) and the politeness of the robot's introduction (polite versus impolite). The physical distance that the participants maintained with the robot in the physical distancing task and the amount of personal information disclosed to the robot in the psychological distancing task were measured as dependent variables.

A total of 60 participants (30 males and 30 females) took part in the experiment. All participants were native English speakers and their ages varied between 18 and 67. The experimenter told the participants that they would be playing a game with the robot and sought informed consent. In the physical distancing task, the participants sat in front of a computer located approximately 20 feet away from the robot. The task started by the robot introducing itself and the task to the participants using either polite or impolite language. In the remainder of the physical distancing task, the experiment software asked the participants to get up and retrieve a word from a list located on the back of the robot a total of five times. When the participants approached the robot, it either looked toward the participants or looked away from the participants. A ceiling camera captured the partic-

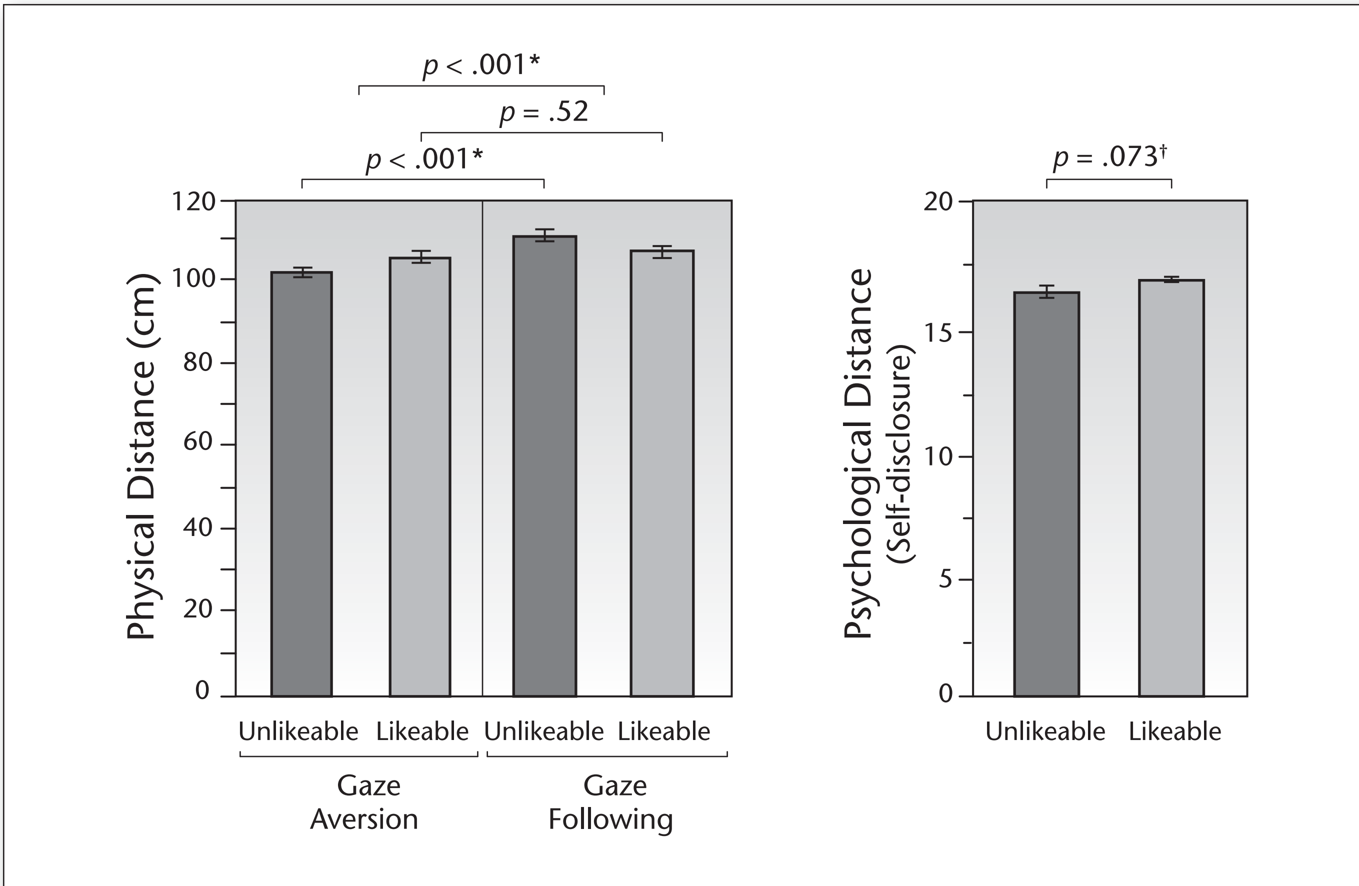

Figure 7. The Participants' Physical Distance and Self-Disclosure Across the Experimental Conditions.

ipants' physical location throughout the task. At the end of the physical distancing task, the participants provided answers to a questionnaire and moved to the psychological distancing task. In the second task, the robot interviewed the participants, asking a total of 17 personal questions. The robot either looked toward the participants or away from the participants during the interview.

The results showed that both manipulations were successful. The participants whom the robot followed with its gaze rated how much the robot looked toward them during the experiment significant higher than those from whom the robot averted its gaze did, $F(1, 58) = 157.40$, $p < .001$. Similarly, the participants with whom the robot used politeness cues rated the robot to be significantly more likable than those who received the impolite introduction did, $F(1, 58) = 7.30$, $p = .009$.

Consistent with the first hypothesis, the participants whom the robot followed with its gaze maintained a larger distance with the robot than those from whom the robot averted its gaze did, $F(1, 584) = 13.66$, $p < .001$. The results also showed a significant interaction between the gaze and rapport manipulations, $F(1, 584) = 7.86$, $p = .005$, providing partial support for the second hypothesis. Of the participants who did not have rapport with the robot, those whom the robot followed with its gaze distanced themselves significantly more than those from whom the robot averted its gaze did, $F(1, 584) = 20.75$, $p < .001$, while the robot's gaze did not affect participants who had rapport with the robot, $F(1, 584) = 0.41$, $p = .52$. Finally, inconsistent with the third hypothesis, the robot's gaze did not affect how much the participants disclosed to the robot, $F(1, 58) = 0.37$, $p = .54$. On the other hand, the amount of rapport participants had with the robot affected disclosure; the participants who had rapport with the robot disclosed marginally more personal information to the robot than those who did not have rapport with the robot did, $F(1, 56) = 3.35$, $p = .073$ (figure 7).

The findings of this study show that, using gaze and politeness cues, robots can shape people's proxemic relationship with them. The gaze cues affected how much physical distance people maintained with the robot, particularly when they did not have rapport with the robot. Politeness cues affected how much rapport people had with the

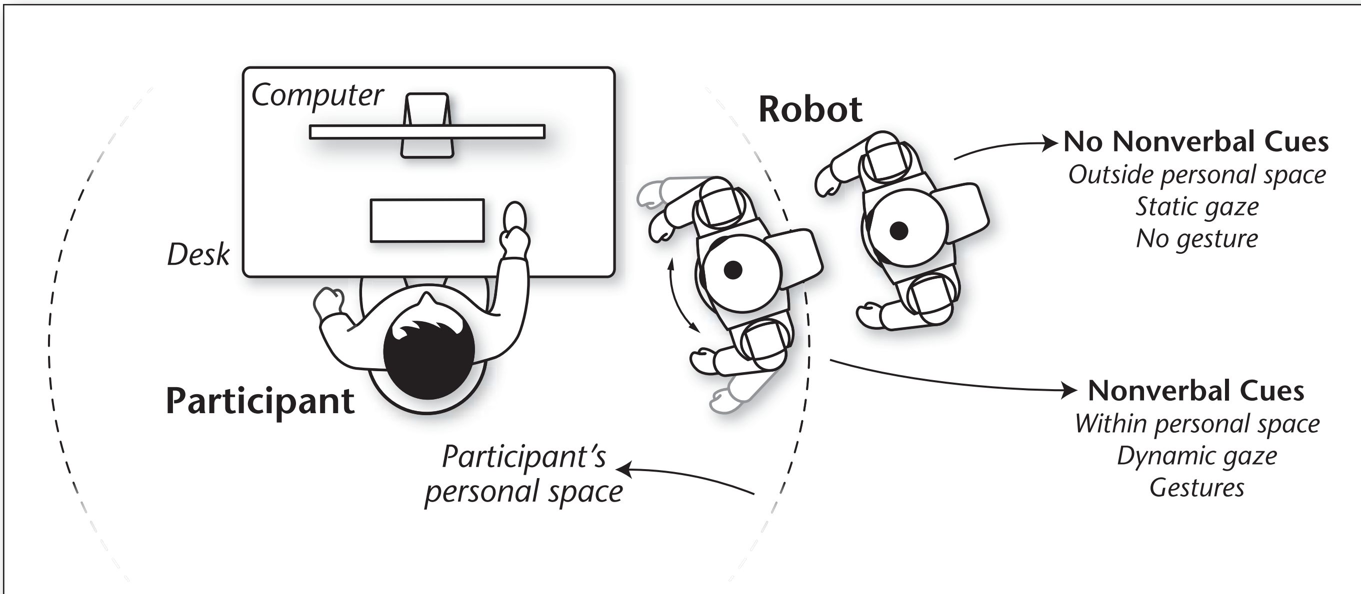

Figure 8. The Setup of the Fourth Study.

robot and how much personal information they disclosed to the robot.

Vocal Cues, Nonverbal Cues, and Persuasion

The last study explored how vocal and nonverbal cues, particularly intonation and the combination of gaze, proximity, and gestures, might affect the persuasiveness of a robot's messages. The hypotheses included three predictions: (1) the persuasiveness of the robot's messages would be higher when the robot uses immediacy cues — vocal, nonverbal, or both — than they would be when it does not use these cues; (2) they would be higher when the robot uses only nonverbal cues than they would be when it uses only vocal cues; and (3) they would be higher for women than they would be for men when the robot uses nonverbal cues.

The study involved a desert survival task in which participants created a ranking of 15 items based on how much these items would increase their chances of surviving in the desert and Wakamaru provided suggestions for changing the rankings of a subset of the items (figure 8). The independent variable was cue type with four levels: (1) no immediacy cue, (2) vocal cues only, (3) nonverbal cues only, and (4) vocal and nonverbal cues. To vary its vocal cues, the robot changed the variability in pitch, creating monotone and expressive versions of its speech. The robot sought to achieve high or low levels of nonverbal immediacy by manipulating its gaze (whether it looked toward the item on the screen when it referred to it in its speech), gestures (whether it used gestures to refer to the item, describe the item's function and importance for survival, and emphasize aspects of its speech), and proximity (whether the robot stayed within or outside the participant's personal space) at the same time. The dependent variables included the number of changes the participants made in their initial rankings based on the robot's suggestions and the participants' evaluations of the social and intellectual characteristics of the robot.

Thirty-two native-English-speaking participants (16 males and 16 females) took part in the study. Their ages ranged between 19 and 49. The study started by seeking informed consent from the participants and providing them with a description of their task. The participants sat in front of a computer where they ranked the desert survival items and reviewed their rankings based on the robot's suggestions. The robot stood next to the participants in a way that it would be visible to the participants and the participants' computer screen would be visible to the robot. After completing the task, the participants filled in a post-experiment questionnaire.

The manipulation checks confirmed that the manipulations in the robot's vocal and nonverbal cues were mostly successful. The participants rated the expressiveness of the robot's voice marginally higher when the robot used higher pitch variability than they did when the robot used low pitch variability, $F(1, 30) = 3.77$, $p = .062$. The participants' ratings of how much the robot used gaze and gesture cues were significantly higher when the robot employed these cues than when it did not, $F(1, 30) = 47.25$, $p < .001$ and $F(1, 30) = 142.08$, $p < .001$ for gaze and gestures, respectively.

The results confirmed the first hypothesis; the participants followed more of the robot's suggestions when the robot employed immediacy cues (vocal, nonverbal, or both) than they did when the robot did not employ these cues, $F(1, 28) = 7.53$, $p = .011$. In particular, they followed more suggestions when the robot employed nonverbal cues than they did when the robot did not employ nonverbal cues, $F(1, 28) = 12.35$, $p = .002$, while the use of vocal cues did not increase the number of suggestions that the participants followed compared to the baseline condition, $F(1, 28) = 1.00$, $p = .32$. Consistent with the second hypothesis, nonverbal cues led participants to follow more of the robot's suggestions than vocal cues did, $F(1, 28) = 6.31$, $p = .018$. Finally, men and women did not differ in how they were affected by immediacy cues, $F(1, 24) = 0.02$, $p = .90$, contradicting with the prediction (figure 9).

The findings of this study confirmed the relationship between immediacy cues and persuasion in human-robot interaction. Immediacy cues improved the persuasiveness of the robot's messages evidenced by the participants' compliance with the suggestions of the robot. Nonverbal cues served as effective immediacy cues, while vocal cues did not achieve the same outcome.

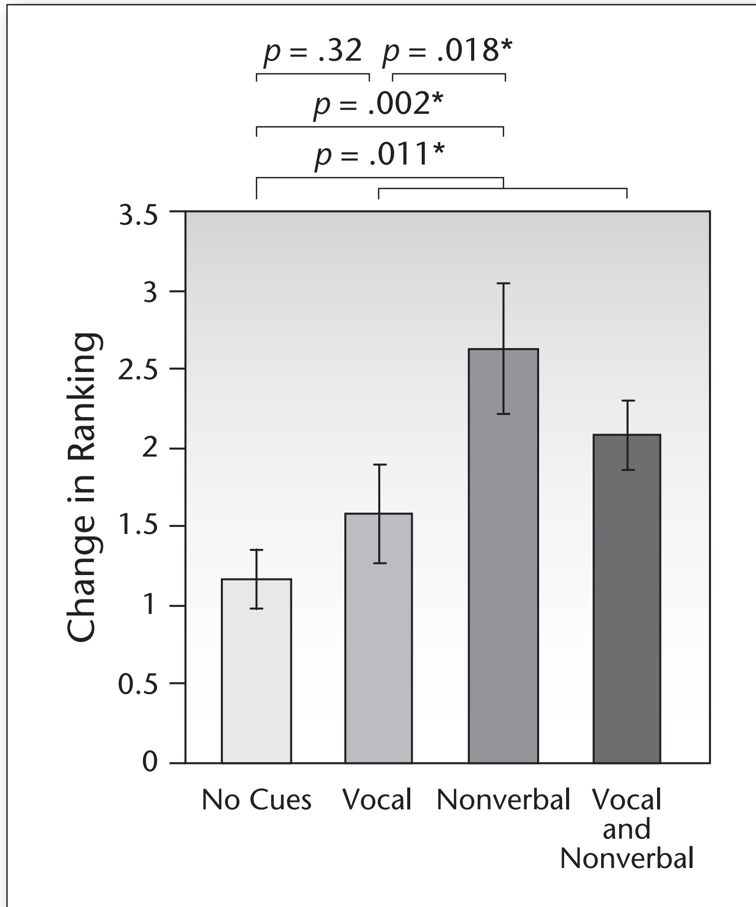

Figure 9. The Number of Items Rankings of Which the Participants Changed Based on the Robot's Suggestions Across the Four Conditions.

Discussion

This article seeks to underline the fundamental connection between embodied communication cues and key social, cognitive, and task outcomes and present this connection as an opportunity for designing embodied dialogue with robots. By gaining a better understanding of how most effective communicators use these cues to achieve such outcomes and using this understanding to carefully tune robots' embodied cues, robot designers might harness the full potential of robotic technology. The research program described here exemplifies how, through carefully designed changes in their embodied cues, robots might elicit such outcomes as learning, rapport, and persuasion. The potential for impact goes beyond these illustrative domains. Building robots that might achieve such outcomes in the real world, however, bears a number of challenges. The paragraphs below discuss some of these challenges and offer potential directions for future research.

Embodied interaction is fundamentally a joint activity in which the embodied cues of all parties in the interaction work together in coordination (Clark 1996; Clark and Brennan 1991; Clark and Wilkes-Gibbs 1986). In the context of human-robot interaction, the control of the embodied cues of the robot necessarily needs to be coordinated with the embodied cues of the users over the course of the interaction. However, most studies of embodied cues in human-robot interaction to date, including to an extent the research program described here, minimally consider the dynamic coordination of communicative cues between the robot and its users. This unilateral approach is largely due to a lack of models that describe coordinated actions in dynamic interaction. Human communication researchers have discussed the limitations of using "single-rate" variables that assume behaviors of individuals to be independent from those of others and proposed, instead, the use of "action-sequence" variables that capture sequences of joint actions in interaction (Duncan et al. 1984). Progress toward overcoming this significant challenge requires a closer look at embodied interaction and the development of specific artificial intelligence techniques suited to model complex joint actions.

The physical, social, and task context of the interaction greatly shapes how parties employ embodied communication cues. For instance, research on the effects of situational factors on gaze behavior has found that the visual complexity of task-relevant objects such as a map and the environment have a significant effect on the gaze behaviors of the participants (Argyle and Graham

1976). Similarly, social factors such as the degree of closeness among parties in interaction (for example, whether the participants are friends or strangers) affect how the parties employ embodied cues such as gaze, gestures, smiling, proximity, and verbal cues and their social outcomes (Stinson and Ickes 1992). Studies of embodied interaction between humans and robots to date have considered such contextual factors only minimally. Most work in this area including the studies presented in the previous section involve controlled, invariable physical, social, and task contexts and lack an exploration of how embodied cues and their outcomes might change across contexts. Future research must seek a better understanding of how various characteristics of context affect embodied interaction and how robots might adapt their use of embodied cues across contexts to maximize their effect in achieving social, cognitive, and task outcomes.

Studies of embodied interaction in human communication also show that parties change their use of embodied cues over the course of the interaction. For instance, the amount of eye contact that parties in a conversation maintain decreases as the interaction unfolds (Abele 1986). The parties in the interaction might also adapt their behaviors to those of their partners; studies of dyadic interaction have shown that members of dyads mimic their partners' nonverbal cues such as facial expressions and smiling (Chartrand and Bargh 1999) and verbal cues such as word choice (Scissors, Gill, and Gergle 2008) and that mimicking the behaviors of a partner improves social outcomes (Chartrand and Bargh 1999). How robots might change their use of embodied cues over the course of an episode of interaction and how they might adapt their behaviors to those of their users remain unexplored. The development of effective embodied cues for human-robot dialogue must explore how the joint use of these cues by humans and robots might evolve and coadapt and what techniques might best model these temporal, interdependent changes in behavior.

Conclusion

Robots are unique among computational systems in their ability to afford embodied interaction. Research in human communication has shown that the cues that compose embodied interaction, when employed effectively, generate significant social, cognitive, and task outcomes from improved learning to increased performance in collaborative work. Robots, with their unique ability to represent and control embodied cues, hold tremendous potential for drawing on this connection to generate similar positive outcomes in human-robot dialogue. Achieving this potential, however, requires a better understanding of the most effective ways in which robots might use embodied cues to generate positive outcomes and the limitations in the extent to which the connection between embodied cues and positive outcomes exists in human-robot interaction. To achieve this understanding, robot designers must systematically study the relationship between particular embodied cues and outcomes, iteratively fine-tuning control parameters for these cues to identify their most effective use to obtain the targeted outcomes.

This article sought to highlight the link between embodied communication cues and significant positive outcomes, provide overviews of embodied cues and outcomes that human communication researchers have considered, and describe a research program that aims to understand the relationship between robot embodied cues and significant positive outcomes in human-robot interaction. The illustrative set of studies from this research program show that, through effective use of verbal, vocal, and nonverbal cues, robots can generate such positive outcomes as improved learning, rapport, and persuasion. While these results are promising, achieving the full potential of embodied human-robot interaction poses further challenges such as the need to better understand the effects of joint activity, contextual factors, time, and adaptation on embodied interaction. More systematic research on embodied human-robot interaction and advancements in these challenges will enable the design of robots that offer people not only representations and interactions with which they are intimately familiar but also significant improvements in key areas of their lives.

Acknowledgments

I would like to thank all collaborators and students who have contributed to the research program described here, particularly Jodi Forlizzi, Jessica Hodgins, Jonathan Mumm, Erin Spannan, Sean Andrist, Vijay Chidambaram, and Yueh-Hsuan Chiang. The coeditors of this special issue, Dan Bohus, Eric Horvitz, Takayuki Kanda, and Antoine Raux, have provided valuable feedback on the earlier versions of this article. This research has been supported by National Science Foundation grants IIS-0121426 and IIS-1017952 and equipment loans from Honda R&D Co., Ltd., and Mitsubishi Heavy Industries, Ltd.

References

Abele, A. 1986. Functions of Gaze in Social Interaction: Communication and Monitoring. *Journal of Nonverbal Behavior* 10(2): 83–101.

Alibali, M., and Nathan, M. 2007. Teachers' Gestures as a Means of Scaffolding Students' Understanding: Evidence from an Early Algebra Lesson. *Video Research in the Learning Sciences*. Mahwah, NJ: Lawrence Erlbaum.

Ambady, N., and Rosenthal, R. 1992. Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. *Psychological Bulletin* 111(2): 256.

Ambady, N.; Koo, J.; Lee, F.; and Rosenthal, R. 1996. More Than Words: Linguistic and Nonlinguistic Politeness in Two Cultures. *Journal of Personality and Social Psychology* 70(5): 996.

Argyle, M., and Graham, J. 1976. The Central Europe Experiment: Looking at Persons and Looking at Objects. *Journal of Nonverbal Behavior* 1(1): 6–16.

Argyle, M. 1988. *Bodily Communication*. Boston, MA: Taylor & Francis.

Baldwin, D. 1995. *Understanding the Link Between Joint Attention and Language*, 131–158. Mahwah, NJ: Lawrence Erlbaum.

Bateson, M.; Nettle, D.; and Roberts, G. 2006. Cues of Being Watched Enhance Cooperation in a Real-World Setting. *Biology Letters* 2(3): 412.

Bayliss, A.; Paul, M.; Cannon, P.; and Tipper, S. 2006. Gaze Cuing and Affective Judgments of Objects: I Like What You Look At. *Psychonomic Bulletin and Review* 13(6): 1061.

Bluethmann, W.; Ambrose, R.; Diftler, M.; Askew, S.; Huber, E.; Goza, M.; Rehnmark, F.; Lovchik, C.; and Magruder, D. 2003. Robonaut: A Robot Designed to Work with Humans in Space. *Autonomous Robots* 14(2): 179–197.

Breazeal, C. 2003. Toward Sociable Robots. *Robotics and Autonomous Systems* 42(3-4): 167–175.

Brennan, S.; Chen, X.; Dickinson, C.; Neider, M.; and Zelinsky, G. 2008. Coordinating Cognition: The Costs and Benefits of Shared Gaze During Collaborative Search. *Cognition* 106(3): 1465–1477.

Brooks, D. 1985. The Teacher's Communicative Competence: The First Day of School. *Theory into Practice* 24(1): 63–70.

Brown, P., and Levinson, S. 1988. *Politeness*. Cambridge: Cambridge University Press.

Brunner, L. 1979. Smiles Can Be Back Channels. *Journal of Personality and Social Psychology* 37(5): 728.

Burgoon, J. 1991. Relational Message Interpretations of Touch, Conversational Distance, and Posture. *Journal of Nonverbal Behavior* 15(4): 233–259.

Burgoon, J.; Birk, T.; and Pfau, M. 1990. Nonverbal Behaviors, Persuasion, and Credibility. *Human Communication Research* 17(1): 140–169.

Burgoon, J.; Bonito, J.; Ramirez, A.; Dunbar, N.; Kam, K.; and Fischer, J. 2002. Testing the Interactivity Principle: Effects of Mediation, Propinquity, and Verbal and Nonverbal Modalities in Interpersonal Interaction. *Journal of Communication* 52(3): 657–676.

Burnham, T., and Hare, B. 2007. Engineering Human Cooperation. *Human Nature* 18(2): 88–108.

Cassell, J. 2001. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine* 22(4): 67.

Chartrand, T., and Bargh, J. 1999. The Chameleon Effect: The Perception–Behavior Link and Social Interaction. *Journal of Personality and Social Psychology* 76(6): 893.

Clark, H. 1996. Using Language. *Computational Linguistics* 23(4): 597–618.

Clark, H., and Brennan, S. 1991. Grounding in Communication. *Perspectives on Socially Shared Cognition* 13(1991): 127–149.

Clark, H., and Marshall, C. 2002. Definite Reference and Mutual Knowledge. *Psycholinguistics: Critical Concepts in Psychology* 414.

Clark, H., and Wilkes-Gibbs, D. 1986. Referring as a Collaborative Process* 1. *Cognition* 22(1): 1–39.

Crusco, A., and Wetzel, C. 1984. The Midas Touch. *Personality and Social Psychology Bulletin* 10(4): 512.

Cutler, A.; Dahan, D.; and Van Donselaar, W. 1997. Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Sspeech* 40(2): 141.

Dautenhahn, K., and Werry, I. 2004. Towards Interactive Robots in Autism Therapy: Background, Motivation, and Challenges. *Pragmatics and Cognition* 12(1): 1–35.

Duncan, S. 1972. Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology* 23(2): 283.

Duncan, S.; Kanki, B.; Mokros, H.; and Fiske, D. 1984. Pseudounilaterality, Simple-Rate Variables, and Other Ills to Which Interaction Research Is Heir. *Journal of Personality and Social Psychology* 46(6): 1335.

Eaton, M.; Mitchell-Bonair, I.; and Friedmann, E. 1986. The Effect of Touch on Nutritional Intake of Chronic Organic Brain Syndrome Patients. *Journal of Gerontology* 41(5): 611.

Ekman, P. 1993. Facial Expression and Emotion. *American Psychologist* 48(4): 384.

Emery, N. 2000. The Eyes Have It: The Neuroethology, Function and Evolution of Social Gaze. *Neuroscience and Biobehavioral Reviews* 24(6): 581–604.

Frischen, A.; Bayliss, A.; and Tipper, S. 2007. Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. *Psychological Bulletin* 133(4): 694.

Fry, R., and Smith, G. 1975. The Effects of Feedback and Eye Contact on Performance of a Digit-Coding Task. *The Journal of Social Psychology*.

Fussell, S.; Setlock, L.; Yang, J.; Ou, J.; Mauer, E.; and Kramer, A. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Human–Computer Interaction* 19(3): 273–309.

Gallace, A., and Spence, C. 2010. The Science of Interpersonal Touch: An Overview. *Neuroscience and Biobehavioral Reviews* 34(2): 246–259.

Goffman, E. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor.

Goffman, E. 1969. *Strategic Interaction*. Philadelphia: University of Pennsylvania Press.

Graham, J., and Heywood, S. 1975. The Effects of Elimination of Hand Gestures and of Verbal Codability on Speech Performance. *European Journal of Social Psychology* 5(2): 189–195.

Guéguen, N. 2004. Nonverbal Encouragement of Participation in a Course: The Effect of Touching. *Social Psychology of Education* 7(1): 89–98.

Hadar, U.; Steiner, T.; Grant, E.; and Rose, F. 1984. The Timing of Shifts of Head Postures During Conservation. *Human Movement Science* 3(3): 237–245.

Haley, K., and Fessler, D. 2005. Nobody's Watching?: Subtle Cues Affect Generosity in an Anonymous Economic Game. *Evolution and Human Behavior* 26(3): 245–256.

Hall, E. 1966. *The Hidden Dimension*. New York: Doubleday.

Hancock, J., and Dunham, P. 2001. Language Use in Computer-Mediated Communication: The Role of Coordination Devices. *Discourse Processes* 31(1): 91–110.

Harris, M., and Rosenthal, R. 2005. No More Teachers' Dirty Looks: Effects of Teacher Nonverbal Behavior on Student Outcomes. *Applications of Nonverbal Communication* 157–192.

Hartelius, E. 2008. *The Rhetoric of Expertise*. Ph.D. Dissertation, University of Texas at Austin, Austin, TX.

Hayduk, L. 1981. The Shape of Personal Space: An Experimental Investigation. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement* 13(1): 87.

Hornik, J. 1992. Tactile Stimulation and Consumer Response. *The Journal of Consumer Research* 19(3): 449–458.

Hornik, J., and Ellis, S. 1988. Strategies to Secure Compliance for a Mall Intercept Interview. *Public Opinion Quarterly* 52(4): 539.

Jones, S., and Yarbrough, A. 1985. A Naturalistic Study of the Meanings of Touch. *Communication Monographs* 52(1): 19–56.

Jourard, S., and Friedman, R. 1970. Experimenter-Subject "Distance" and Self-Disclosure. *Journal of Personality and Social Psychology* 15(3): 278.

Kanda, T.; Hirano, T.; Eaton, D.; and Ishiguro, H. 2004. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human–Computer Interaction* 19(1): 61–84.

Kendon, A. 1973. *The Role of Visible Behaviour in the Organization of Social Interaction 1*, volume 4. 29–74. Boston, MA: Academic Press.

Kirk, D.; Rodden, T.; and Fraser, D. 2007. Turn It This Way: Grounding Collaborative Action with Remote Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1039–1048. New York: Association for Computing Machinery.

Krauss, R. 1998. Why Do We Gesture When We Speak? *Current Directions in Psychological Science* 7(2): 54–60.

Kuhl, P. 2007. Is Speech Learning "Gated" by the Social Brain? *Developmental Science* 10(1): 110–120.

Liu, C.; Kay, D.; and Chai, J. 2011. Awareness of Partners Eye Gaze in Situated Referential Grounding: An Empirical Study. In *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction*. New York: Association for Computing Machinery.

Mason, M.; Tatkow, E.; and Macrae, C. 2005. The Look of Love. *Psychological Science* 16(3): 236.

McClave, E. 2000. Linguistic Functions of Head Movements in the Context of Speech. *Journal of Pragmatics* 32(7): 855–878.

McNeill, D. 2005. *Gesture and Thought*. Chicago, IL: University of Chicago Press.

Mehrabian, A. 1966. Immediacy: An Indicator of Attitudes in Linguistic Communication. *Journal of Personality* 34(1): 26–34.

Mehrabian, A. 1967. Attitudes Inferred From Neutral Verbal Communications. *Journal of Consulting Psychology* 31(4): 414.

Mehrabian, A. 1969. Significance of Posture and Position in the Communication of Attitude and Status Relationships. *Psychological Bulletin* 71(5): 359.

Mehrabian, A. 1972. A Nonverbal Communication. Chicago, IL: Aldine Atherton.

Montagu, A., and Matson, F. 1979. *The Human Connection*. McGraw-Hill

Morales, M.; Mundy, P.; Delgado, C.; Yale, M.; Messinger, D.; Neal, R.; and Schwartz, H. 2000. Responding to Joint Attention Across the 6 Through 24 Month Age Period and Early Language Acquisition. *Journal of Applied Developmental Psychology* 21(3): 283–298.

Mutlu, B.; Forlizzi, J.; and Hodgins, J. 2006. A Storytelling Robot: Modeling and Evaluation of Human-Like Gaze Behavior. In *Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots*, 518–523. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Osborn, D. 1996. Beauty Is as Beauty Does?: Makeup and Posture Effects on Physical Attractiveness Judgments. *Journal of Applied Social Psychology* 26(1): 31–51.

Otta, E.; Lira, B.; Delevati, N.; Cesar, O.; and Pires, C. 1994. The Effect of Smiling and of Head Tilting on Person Perception. *The Journal of Psychology* 128(3): 323–331.

Otteson, J., and Otteson, C. 1980. Effect of Teacher's Gaze on Children's Story Recall. *Perceptual and Motor Skills*.

Poggi, I.; D'Errico, F.; and Vincze, L. 2010. Types of Nods. The Polysemy of a Social Signal. In *Proceedings of International Conference on Language Resources and Evaluation*.

Richardson, D., and Dale, R. 2005. Looking to Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science* 29(6): 1045–1060.

Richmond, V.; Gorham, J.; and McCroskey, J. 1987. The Relationship Between Selected Immediacy Behaviors and Cognitive Learning. *Communication Yearbook* 10(574–590).

Russell, J.; Bachorowski, J.; and Fernández-Dols, J. 2003. Facial and Vocal Expressions of Emotion. *Annual Review of Psychology* 54(1): 329–349. Palo Alto, CA: Annual Reviews, Inc.

Scherer, K.; Ladd, D.; and Silverman, K. 1984. Vocal Cues to Speaker Affect: Testing Two Models. *Journal of the Acoustical Society of America* 76(5): 1346–1356.

Scissors, L.; Gill, A.; and Gergle, D. 2008. Linguistic Mimicry and Trust in Text-Based CMB. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 277–280. New York: Association for Computing Machinery.

Segrin, C. 1993. The Effects of Nonverbal Behavior on Outcomes of Compliance Gaining Attempts. *Communication Studies* 44(3): 169–187.

Sherwood, J. 1987. Facilitative Effects of Gaze upon Learning. *Perceptual and Motor Skills*.

Stinson, L., and Ickes, W. 1992. Empathic Accuracy in the Interactions of Male Friends Versus Male Strangers. *Journal of Personality and Social Psychology* 62(5): 787.

Thompson, L., and Massaro, D. W. 1994. Children's Integration of Speech and Pointing Gestures in Comprehension. *Journal of Experimental Child Psychology* 57(3): 327–354.

Ullmer, B., and Ishii, H. 2000. Emerging Frameworks for Tangible User Interfaces. *IBM Systems Journal* 39(3.4): 915–931.

Van Houten, R.; Nau, P.; MacKenzie-Keating, S.; Sameoto, D.; and Colavecchia, B. 1982. An Analysis of Some Variables Influencing the Effectiveness of Reprimands. *Journal of Applied Behavior Analysis* 15(1): 65.

Wennerstrom, A. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis*. New York: Oxford University Press.

Woolfolk, A., and Brooks, D. 1985. The Influence of Teachers' Nonverbal Behaviors on Students' Perceptions and Performance. *The Elementary School Journal* 85(4): 513–528.

Bilge Mutlu is an assistant professor of computer science, psychology, and industrial engineering at the University of Wisconsin–Madison where he directs a research program on designing social robots and their applications in education, health and well-being, and collaborative work. He received his Ph.D. from Carnegie Mellon University's Human-Computer Interaction Institute in 2009.