Designing Gaze Behavior for Humanlike Robots

Bilge Mutlu

May 2009

CMU-HCII-09-101

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

Thesis Committee:
Jodi Forlizzi (Co-chair)
Jessica Hodgins (Co-chair)
Sara Kiesler
Justine Cassell (Northwestern University)

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

This work is funded in part by National Science Foundation grants ITR-IIS-0121426, HSD-IIS-0624275, and CRI-CNS-0709077; JSPS Grant-in-Aid for Scientific Research (S), KAKENHI (20220002); and fellowships and equipment grants by ATR International, Ford Motor Company, Honda R&D Co., Ltd., and Mitsubishi Heavy Industries, Ltd. Any opinions, findings, or recommendations expressed in this material are those of the author and do not necessarily reflect those of these funding agencies.

Keywords

Gaze, social gaze, human-robot interaction, human-computer interaction, human factors, social robots, humanlike robots, robotics, ASIMO, Robovie, Geminoid, modeling human behavior, empirical studies, discourse analysis, language use, computational modeling, oratory, oratorial gaze, attention, information recall, interaction design, person perception, positive evaluation, conversations, conversational gaze, footing, participation structure, participant roles, turn-taking, gaze patterns, information structure, Japanese language, feelings of groupness, liking, nonverbal leakage, gaze cues, gaze cueing, mental states, attentional states, task performance, social desirability, and pet ownership.

Abstract

Humanlike robots are designed to communicate with people using human verbal and nonverbal language. Social gaze cues play an important role in this communication. Although research in human-robot interaction has shown that people understand these gaze cues and interpret them as valid signals for human communication, whether they can serve as effective communicative mechanisms and lead to significant social and cognitive outcomes in human-robot interaction remains unknown. Furthermore, the theoretical and empirical foundations for how these mechanisms might be designed to work with the human communicative system have not been systematically explored.

The research questions that I seek to address in this area are as follows: What are the design variables for social gaze behavior? How do we design gaze behavior for humanlike robots? Can designed behaviors lead to positive, significant social and cognitive outcomes in human-robot interaction such as better learning, stronger affiliation, and active participation in conversations?

This dissertation seeks to find answers to these questions by exploring the design space to identify design variables for social gaze, adapting an approach based on modeling human behavior to designing robot behaviors, and evaluating the social and cognitive outcome of designed behaviors in three studies that focus on different functions of social gaze behavior using three robotic platforms, ASIMO, Robovie, and

Geminoid. The first study focused on designing gaze behavior for communication of attention and found strong learning effects induced by a simple manipulation of how much ASIMO looks at an individual. The second study looked at how Robovie might use gaze cues to shape the participant roles of its conversational partners and found strong effects of gaze cues in behavioral and subjective measures of participation, attentiveness, liking, and feelings of groupness. The final study, which consisted of three experiments, explored how Robovie and Geminoid might use gaze cues to communicate mental states, and found task performance effects in a guessing game led by attributions of mental states.

This research contributes to the design of robotic systems with a theoretically and empirically grounded methodology for designing communicative mechanisms, human-robot interaction research with a better understanding of the social and cognitive outcomes of interacting with robots, and human communication research with new knowledge on and computational models of human gaze mechanisms.

Acknowledgements

First and foremost, I would like to thank my co-advisors, Jodi Forlizzi and Jessica Hodgins for joining forces to help me with the innumerable facets of my work and mentor me towards becoming an academic. I am incredibly grateful to Sara Kiesler for her guidance throughout my doctoral education and for serving in my committee. I would also like to thank my external committee member, Justine Cassell, for providing me with advice, feedback, and inspiration during this research.

My family deserves special thanks for their continuing love, support, and encouragement, including my fiancée Lindsay Jacobs who carefully read my dissertation from cover to cover, my brother Erdem Ilker, and my parents Hanife and Mustafa.

I am thankful to all my co-authors and collaborators for their hard work and thoughtful, valuable contributions including Susan Fussell, Norohiro Hagita, Hiroshi Ishiguro, Takayuki Kanda, Toshiyuki Shiwa, and Fumitaka Yamaoka.

Several others outside Carnegie Mellon University have provided help and contributed to discussions around this work including Jeremy Bailenson, Nathan Frier, Peter Kahn, Clifford Nass, and Leila Takayama, as well as organizers and participants of DIS 2006 and CHI 2008 Doctoral Consortiums.

I also offer many thanks to several students, faculty, staff, and friends that made this work possible through their support, ideas, and friendship, including Turadg Aleahmad, Sonya Allin, Lisa Anthony, Chris Atkeson, Daniel Avrahami, Aruna Balakrishnan, Ryan Baker, Aaron Bauer, Jo Bodnar, Moira Burke, Melissa Carrozza, David Casillas, Marian D'Amico, Laura Dabbish, Scott Davidoff, Rob DeLine, Anind Dey, Tawanna Dillahunt, Carl DiSalvo, Matthew Easterday, Frederick Eberhardt, James Fogerty, Darren Gergle, Rachel Gockley, Carlos Guestrin, Gahgene Gweon, Elspeth Golden, Chris Harrison, Gary Hsieh, Jason Hong, Scott Hudson, Amy Hurst, Bonnie John, Ken Koedinger, Andreas Krause, Robert Kraut, Queenie Kravitz, Johnny Lee, Joonhwan Lee, Min Kyung Lee, Ian Li, Marti Louw, Jeanne McDade, Marek Michalowski, Philipp Michel, Brad Myers, Jeff Nichols, Christine Neuwirth, Illah Nourbakhsh, Sue O'Conner, Eric Paulos, Randy Pausch, Minoli Ratnatunga, Brandy Renduels, Ido Roll, Carolyn Rosé, Aubrey Shick, Mary Scott, Peter Scupelli, Fleming Seay, Kazuhiko Shinozawa, Dan Siewiorek, Reid Simmons, Karen Tang, Cristen Torrey, Jennifer Turken, Erin Walker, Jason Weise, Nicole Willis, Jake Wobbrock, Jeff Wong, Ruth Wylie, and John Zimmerman.

Thank you, also, to the corporate and governmental sponsors of this work, including National Science Foundation, Japan Society for the Promotion of Science, ATR International, Honda R&D Co., Ltd., Ford Motor Company, and Mitsubishi Heavy Industries, Ltd.

To my family,

Hanife, *Mustafa*, and *Erdem Ilker Mutlu*,

for a lifetime of love and support.

Table of Contents

List of Figures

List of Tables

1. Introduction

In the future, humanlike robots might serve as informational agents in public spaces, as caregivers or companions for the elderly, and as educational peers for children. These and other service tasks will require that robots communicate using human verbal and nonverbal language and carry out conversations with people. In these tasks, gaze will play an important role. For example, suppose that an educational robot's task is to tell stories at a primary school and make sure that everyone in the class is following the story. What would the robot do if it realized that one of the students were not attending to its story? What would human teachers do? The following excerpt provides some insight into these questions (Woolfolk & Brooks, 1985):

Professor:	*How do you know when your teacher really means what she says?*
Third Grader:	*Well,* ***her eyes get big*** *and round and she* ***looks right at us****. She doesn't move and her voice is a little louder, but she talks kinda slowly. Sometimes she stands over us and* ***looks down at us****.*
Professor:	*What happens then?*
Third Grader:	*The class does what she wants!*

As in the excerpt above, human teachers change aspects of their verbal and nonverbal language—particularly gaze, as highlighted in bold—to communicate to their students that they should be attending to the teacher. In fact, research has shown that simply looking at that student will improve learning (Otteson & Otteson, 1980; Sherwood, 1987). What should the robot in our scenario do? The apparent solution is for the robot to look at the distracted student more. However, whether robots can use human communicative mechanisms to evoke social and cognitive outcomes in people such as improved attention or learning is unknown.

Researchers have been developing robotic systems that are designed to support human communicative mechanisms for nearly a decade (Breazeal, 1998; Brooks et al., 1999; Nourbakhsh, 1999; Scassellati, 2001; Dautenhahn et al., 2002; Kanda et al., 2002; Pineau et al., 2003; Minato et al., 2004). A number of studies have shown the importance of gaze behavior in human-robot communication (Imai et al., 2002; Sidner et al., 2004; Yoshikawa et al., 2006; Yamazaki et al., 2008). For example, Imai et al. (2002) showed that people can accurately interpret a robot's orientation of attention using cues from its gaze. When the robot's gaze behavior was contingent with that of participants, people had stronger "feelings of being looked at" (Yoshikawa et al., 2006). In a study by Sidner et al. (2004), the robot's use of gaze cues and gestures significantly increased people's engagement as well as their use of gaze cues to communicate with the robot. Yamazaki et al. (2008) showed that when a robot followed simple rules of conversational turn-taking to coordinate its gaze behavior and verbal utterances, people were more likely to display nonverbal behaviors at turn boundaries.

Although these studies provide some evidence that robot gaze affects people's behavior, a systematic study of how gaze could lead to significant social and cognitive outcomes in different situations is still lacking. The following questions remain unanswered; Can robot gaze affect human learning? Can a robot use gaze cues to regulate turn-taking and conversational participation? Can robot gaze help people infer the mental states of the robot? Furthermore, how social gaze behavior should be designed for robots to work with human communicative mechanisms needs further exploration.

This dissertation addresses these questions through developing (1) a methodology for applying human communication patterns to the design of social behaviors for humanlike robots, (2) a set of design variables or behavioral parameters—such as gaze target, frequency, and duration—that designers can use to create gaze behaviors for robots that could be manipulated to obtain social and cognitive outcomes, and (3) a theoretical framework for understanding how robot gaze might serve as a communicative mechanism. This thesis contributes to the design of robotic systems a theoretically and empirically grounded methodology for the design of communicative mechanisms for robots. It also contributes to human-robot interaction research a better understanding of the social and cognitive outcomes of interacting with robots. Finally, it contributes to human communication research new knowledge and computational models of human gaze mechanisms, and a deeper understanding of how human communicative mechanisms respond to artificially created social stimuli.

This chapter describes the robotic platforms used for the studies in this dissertation, the research context that motivates the research questions, and the approach taken for addressing these questions. Chapter 2 provides a review of related work on social gaze from literature on human communication research, human-computer interaction, and robotics, with a specific focus on the functions of gaze considered in this dissertation. Chapters 3 to 5 provide details on the design of and results from three empirical studies that focused on three functions of gaze: communication of a speaker's attention, regulation of conversational roles in triads, and communication of a speaker's mental states. Chapter 6 presents some of the limitations of this work and

Figure 1.1. The three robotic platforms used for the studies in this dissertation: Honda's ASIMO (left), ATR's Robovie R-2 (middle), and ATR's Geminoid (right).

how future research might address these limitations. Finally, Chapter 7 lists the conclusions and contributions of this work.

1.1. Research Platforms

Three robotic platforms were used for the empirical studies in this dissertation (Figure 1.1). Honda's ASIMO (Sakagami et al., 2002) was used for the first study. ASIMO's gaze capabilities include a two-degree-of-freedom head with fixed, black eye spots covered by a transparent shield. In the second study, ATR's Robovie R-2 (Ishiguro et al., 2001) was used. Robovie's gaze capabilities include a three-degree-of-freedom head and independently moving, two-degree-of-freedom eyes, each representing an abstraction of a black iris surrounded by white sclera. Finally, two robots, Robovie and ATR's Geminoid (Nishio et al., 2007), were used in the third study. Geminoid's gaze capabilities include a four-degree-of-freedom head and independently moving, two-degree-of-freedom eyes constructed to provide a realistic representation of the human eye. All three robots provided application programming interfaces (API) that allowed for precise and real-time control of gaze behaviors in degrees and speed of rotation for each degree of freedom.

1.2. Research Context

Most research in robotics builds on a future vision for everyday use of humanoid companions and assistants. Accordingly, the research questions posed in this dissertation are motivated by a set of three future scenarios. They provide context for the three empirical studies that look at how robot gaze might serve as a communicative mechanism and for a methodological inquiry into designing humanlike behavior.

1.2.1. Scenario 1

Jeremy works at the Liberty Elementary School in Pittsburgh, Pennsylvania as an English instructor. ASIMO (Sakagami et al., 2002) is used at this school as an aide to English and history instructors. Jeremy teaches English to third graders and has three classes a week—on Mondays, Wednesdays, and Fridays. On Mondays, ASIMO tells the class stories of Jeremy's choice. On Wednesdays, Jeremy discusses the story with the class and asks the class to write a one-page review of the story and bring it to class on Friday.

Recently, Jeremy has realized that Chloe, one the students in his third grade class, has not been participating in the discussions and her essays are very brief. He talks to Chloe and has a phone conversation with her mother to see if there is any trouble at home. But nothing seems to stand out. He talks to the history and math teachers about the recent change in Chloe's attention, but neither instructor seems to notice a change.

> *Jeremy decides that Chloe might be distracted, or she might even be losing interest in English. He tells ASIMO to pay particular attention to Chloe during class. He hopes to monitor Chloe's behavior and direct her attention to class.*

Jeremy's problem is not uncommon. In fact, research in educational psychology suggests that classroom inattentiveness might have negative effects on literacy (Rowe & Rowe, 1999). However, teachers can positively affect student attentiveness using aspects of nonverbal language such as interpersonal space, gestures, gaze, and tone of voice (Woolfolk & Brooks, 1985). Gaze being directed at students, in particular, is shown to improve learning in primary school children (Otteson & Otteson, 1980) and college students (Sherwood, 1987). Could these results transfer to robots? If so, then ASIMO should simply look at Chloe more frequently to direct her attention to class.

Researchers have developed pedagogical virtual agents that direct students' attention using gaze cues and gestures (Rickel & Johnson, 1999; Ryokai et al., 2003). The use of gaze cues by robots is also shown to have a positive effect on engagement (Bruce et al., 2002; Sidner et al., 2004). However, whether cues from the gaze of a robot can direct attention in a way that it leads to better learning is yet unknown. Furthermore, how these cues could be designed to provide social and cognitive benefits has not been systematically studied.

The following questions remain unanswered: Can robot gaze communicate attention and lead to better learning? How can we design robot gaze behavior to attract attention and improve learning? What might the design variables be? The first study sought answers to these questions through modeling the gaze behaviors of a human storyteller, creating gaze cues for ASIMO to perform storytelling, and evaluating whether increased gaze would lead participants to have better recall of the robot's story. This study is described in Chapter 3.

1.2.2. Scenario 2

> *Aiko is a shopper at the Namba Parks shopping mall in Osaka where Robovie (Ishiguro et al., 2001) serves as an information booth attendant. Aiko is trying to find the closest Muji store and also wants to know if the store also sells furniture. She approaches Robovie to inquire about the shop.*

The conversational situation that Robovie will have to manage in this scenario is a two-party conversation in which Robovie and Aiko take turns playing the roles of speaker and addressee (Clark, 1996).

> *As Aiko receives information from Robovie about how to get to the Muji store, another shopper, Yukio, approaches Robovie's booth. Yukio wants to get a program of this month's shows at the amphitheater. When Yukio approaches the information booth, Robovie acknowledges Yukio's presence with a short glance but turns back to Aiko signaling to Yukio that Yukio will have to wait until the conversation with Aiko is over.*

What is different in this conversational situation form the previous one is the addition of a non-participant (Clark, 1996) who plays the role of a bystander (Goffman, 1979).

> *After Robovie's conversation with Yukio is over, a couple, Katsu and Mari, approach the booth inquiring about the Korean restaurants in the mall. Robovie asks Katsu and Mari a few questions on their food preferences and—understanding that they don't like spicy food—leads the couple to Shijan located on the sixth floor of the mall.*

This last situation portrays a three-party conversation where Robovie plays the role of the speaker and Katsu and Mari are addressees for most of the conversation. Although Robovie needs to carry on conversations in all of these situations, the differences in its partners' levels of participation require him to provide appropriate social signals to regulate each person's conversational role. When Yukio approaches the booth, Robovie has to make sure that Aiko's status as addressee doesn't change, but also that Yukio's presence is acknowledged and approved. In talking to Katsu and Mari, the robot has to make sure that both feel equally respected as addressees.

Considerable evidence suggests that people use gaze cues to perform this social-regulative behavior (Bales et al., 1951; Schegloff, 1968; Sacks et al., 1974; Goodwin, 1981). Research in human-computer interaction has shown that these cues are also effective in regulating conversational participation when they are used by virtual agents (Bailenson et al., 2005; Rehm & Andre, 2005). Robot gaze is shown to be effective in performing conversational functions such as supporting turn-taking behavior (Kuno et al., 2007; Yamazaki et al., 2008) and showing appropriate listening behavior (Trafton et al., 2008), but how these cues might shape different forms of participation remains unexplored. Furthermore, whether the cues used by humans can be carried over to robots and create social and cognitive outcomes that can be predicted by our knowledge of human communication is unknown.

The second study attempted to answer the following questions: Can simple cues from a robot's gaze lead to different forms of conversational participation? How can we design gaze behavior that leads to such outcomes? What might the design variables be? In the study, the conversational gaze mechanisms of a human speaker were modeled in different participation structures, these mechanisms were created for Robovie, and whether people conformed to the conversational roles that the robot signaled to them and how conforming to these roles affected their experience and evaluations of the robot were experimentally evaluated. This study is described in Chapter 4.

1.2.3. Scenario 3

> *Akira has recently moved to the Osaka area with his four-year-old son, Ken. Because Ken is an only child, and Akira doesn't have many friends with kids in this new town, he has been thinking about enrolling Ken in preschool where he can socialize with other kids. He consults with their new family doctor, Hiromi, during a regular visit about whether she has any recommendations. Hiromi tells Akira that she had recently heard from a child psychologist colleague of a new preschool in the area that focuses on social development. Following Hiromi's suggestion, Akira visits the school and—having liked the school's focus and program very much—decides to enroll Ken.*
>
> *The school uses a variety of methods to facilitate children's social development including the use of a number of interactive technologies. One of the school's programs uses an educational guessing game played with a humanlike robot that is carefully designed to facilitate the development of the ability to read nonverbal cues and make inferences on the mental and emotional states of a partner. Ken starts playing this game with Geminoid, an android robot developed to look extremely humanlike and produce subtle social cues (Nishio et al., 2007).*

In interpreting others' feelings and intentions, we rely not only on explicit and deliberate communicative acts, but also on implicit, seemingly automatic, and unconscious nonverbal cues. When we see the trembling hands of a public speaker, we assume that the speaker is nervous. Similarly, when we suspect that someone might be lying, we look for cues in their nonverbal behavior that would reveal the person's emotional or intellectual state. These examples illustrate a set of behaviors called "nonverbal leakage" cues that are products of internal, cognitive processes and reveal information to others about the mental and emotional states of an individual

(Ekman & Friesen, 1969; Zuckerman et al., 1981). Could Geminoid gradually employ these cues to help facilitate Ken's development of the ability to use nonverbal information to interpret the mental states of a partner?

Research in human communication has shown that naïve observers can identify deception (Ekman & Friesen, 1969; DePaulo et al., 2003), dissembling (Feldman et al., 1978), genuineness of smiles (Surakka & Hietanen, 1998; Williams et al., 2001), friendliness and hostility (Argyle et al., 1971), affective states (Scherer et al., 1972; Scherer et al., 1973; Waxer, 1977; Krauss et al., 1981), and disfluency of speech (Chawla & Krauss, 1994) using nonverbal cues. Furthermore, these behaviors might play an important role in forming impressions of others—a process in which people rely heavily on nonverbal behavior (Ambady & Rosenthal, 1992). Nonverbal leakage cues and, more broadly, seemingly unintentional and non-semantic nonverbal behaviors might pose an important area of inquiry for research in human-robot interaction. Furthermore, the communicative richness of these cues offers opportunities for designing richer and more natural behaviors for robots.

While research in human-robot interaction has made significant progress in understanding the use of cues such as communication of primary emotions through facial expressions (Breazeal & Scassellati, 1999; Scheeff et al., 2000; Breazeal, 2002; Bartneck et al., 2004; Miwa et al., 2004; Gockley et al., 2006), arm and bodily gestures (Tojo et al., 2000), and vocal tone (Breazeal, 2001), how implicit, non-strategic, and non-semantic cues might be used in human-robot communication has not been explored.

The third study of this dissertation sought to find answers to the following questions: Do people detect nonverbal leakages in a robot? If so, do they interpret these messages correctly to attribute mental states to the robot? How do the physical characteristics of the robot affect these inferences? How can we design such subtle cues? This study attempted to answer these questions by gaining an understanding of how people leak information through gaze cues in the context of a guessing game, designing these cues for two robots, Robovie and Geminoid, and evaluating whether people used these cues to make attributions of mental states to the robots. The third study is described in Chapter 5.

1.3. Design Approach

The questions motivated by these scenarios point to a highly complex design problem: How can we design social behavior? Design problems at this level of complexity are often characterized as "wicked" (Rittel & Webber, 1973) or "messy" (Schön, 1983) problems that require the designer to consider an infinite number of parameters that would lead to an infinite number of unique design solutions. The goal of the designer is to make sense of this "design complexity" (Stolterman, 2008) and produce solutions to the design problem that are optimal under given constraints. In doing so, the designer needs to follow a disciplined and rigorous process in which the designer uses the tools, methods, and knowledge required to make sense of the complexity in the design problem (Wolf et al., 2006). In addition to using these resources, the designer "has to make all kinds of decisions and judgments, such as, how to frame the situation,...what to pay attention to, what to dismiss, and how to explore, extract, recognize, and choose useful information from all of these potential sources," making "the designer's judgment... the primary 'tool' in dealing with design complexity" (Stolterman, 2008). However, the designer can also further formalize these intuitions using empirical methods and ground them in existing scientific or practical knowledge within the limits of available resources. In summary, the designer follows an "analytic, partly formalizable, partly empirical, teachable… design process" (Simon, 1996) to create a design "with a specific purpose, for a specific situation, for a specific client and user, with specific functions and characteristics, and done within a limited time and with limited resources" (Stolterman, 2008).

To design social gaze behaviors for the situations described in the three scenarios presented, I developed and followed a process in which I employed methods and knowledge from a number of scientific disciplines, made a number of design decisions in how I used these methods and what variables I focused on, and tried to further formalize these decisions using empirical methods and ground them in existing scientific or practical knowledge. My main criterion for success was obtaining particular measurable social and cognitive benefits. For instance, the second study involved gaining an understanding of a number of conversational gaze mechanisms. In doing so, I employed methods mainly from sociolinguistics to conduct a detailed analysis of the relationship between gaze behavior and interactional processes, following a rigorous scientific process. In this process, I also made a number of design decisions, such as choosing a particular granularity in data analysis. However, I strived to formalize decisions such as this one by grounding them in theory in conversational organization. I evaluated the success of the design of these conversational gaze mechanisms by evaluating whether they led to social and cognitive benefits such as higher conversational participation, more liking, and stronger feelings of groupness.

A graphical illustration of my design process is provided in Figure 1.2. Because finding solutions for messy problems requires the designer to obtain a deep understanding of the social situation of the problem, this process starts with studying gaze behavior in the social context that motivated the design problem in order to identify relevant gaze

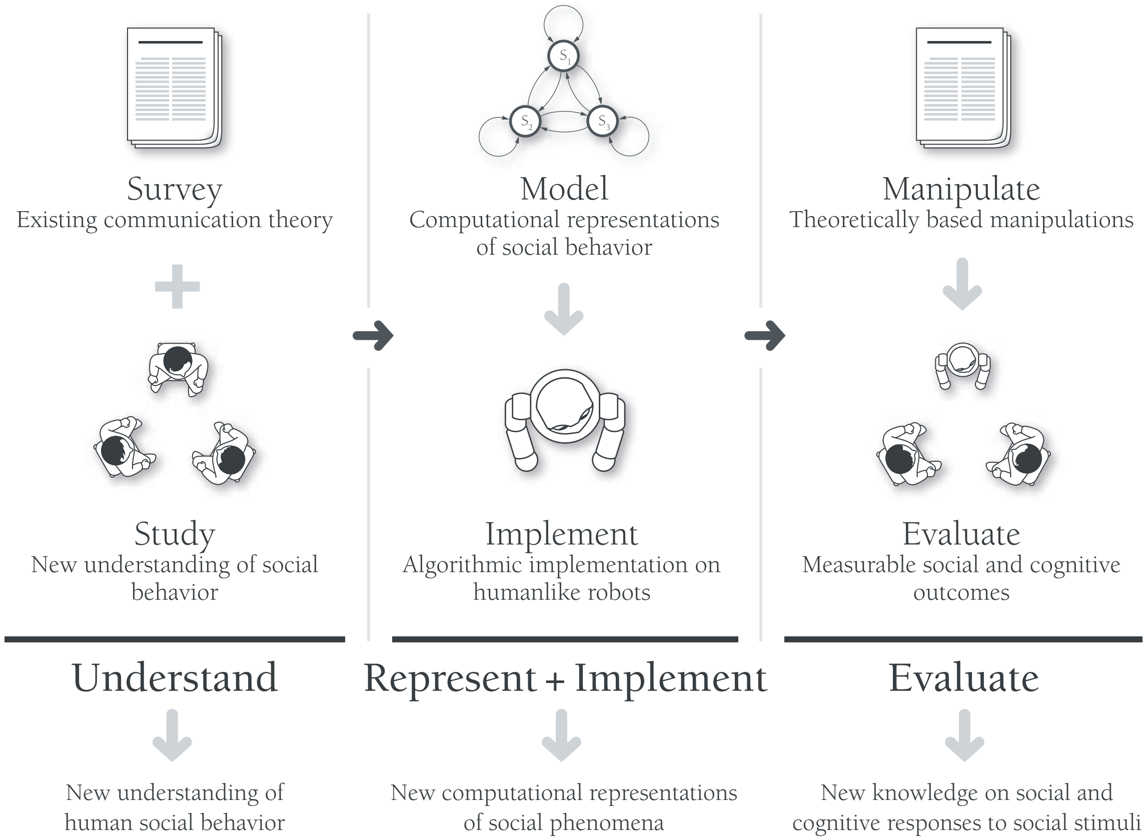


Figure 1.2. The three-stage process followed in this dissertation of understanding, representation and implementation, and evaluation for designing social behavior for humanlike robots.

mechanisms and surveying theories from communication, social and developmental psychology, and sociolinguistics that might explain and predict how these mechanisms function in social interaction. This stage is followed by formalizing this understanding in computational models using knowledge and methods from computational linguistics and implement these models on robotic platforms as computer algorithms using methods from computer science and robotics. Finally, the social and cognitive outcomes of the designed behavior is evaluated through testing theoretically based hypotheses in controlled laboratory experiments in human-robot interaction scenarios using human subjects.

1.3.1. Theoretically and Empirically Grounded Design

The goal of this stage is to ground the process of designing social behavior for humanlike robots in human communication theory and empirical data collected in the social context of interest. In this process, theory provides a top-down, predictive framework for the designed social behavior. For instance, Wang et al. (2005) used politeness theory (Brown & Levinson, 1987) as the basis for their design of a pedagogical agent's social behaviors. Similarly, Yamazaki et al. (2008) used theory on conversational turn-taking (Sacks et al., 1974) to design conversational behaviors for a robot. However, I argue that because behavioral theory is created with the goal of understanding the significance of human behavior in social interaction and representing this significance as predictive, top-down models, theories under-specify the bottom-up elements that are necessary for designing social behavior. For instance, theory on gaze behavior suggests that people occasionally look away from their partners during conversation, but does not describe where they look when they look away, which is an important variable that needs to be considered in designing conversational gaze behavior. Therefore, in this research, I follow a process in which I try to augment these top-down models with bottom-up design specifications that are extracted from empirical data, such as a distribution of how long a speaker looks at a particular target during speech. Examples of this modeling-based design approach also exist; Cassell et al. (2007) collected data from 28 individuals to build a model of how people use gestures to make spatial references in giving direction and used their model to design gestures for a conversational agent called NUMACK. Similarly, Kanda et al. (2007) used data from 25 dyads to model the delay between one's deictic gesture and the other's changing orientation toward the direction of the reference and designed orienting behaviors for a robot based on this model.

Ensuring the validity and generalizability of these models is a major challenge, mainly because of the complexity of and individual differences in social behavior. One approach to addressing this challenge is to use data from a large sample, in examples such as the those described above, to account for individual differences. However, because existing methods for modeling social behavior mainly uses human coders, a detailed analysis of data from a large sample is costly. An alternative approach is to choose a small sample of individuals carefully and analyze more features within this sample of individuals in detail. The research presented in this dissertation follows the latter approach. For example, the second study collected data on conversational gaze behavior from four triads, assessed the data from each triad for conversational fluency, and used the data that showed the highest level of fluency for a detailed analysis. While using gaze data from relatively few individuals imposes some limitations on the generalizability of the gaze models developed in this research, it allowed for conducting a detailed analysis of a variety of communicative gaze mechanisms within each observation. Also, a model developed after a single individual might better capture certain behavioral characteristics that might be lost in an average model developed after a large sample. These and other limitations of the behavioral modeling process are further considered in Chapter 6.

The research presented in this dissertation draws on theories of social gaze behavior from a broad set of areas, particularly sociology, social and cognitive psychology, communication, linguistics, sociolinguistics, and neurophysiology. It uses methods primarily from behavioral and discourse analysis to augment these theories with findings from empirical data. Data collection is done using audio and video recordings using multiple cameras and coding for behavioral variables using human coders.

1.3.2. Representation & Implementation

The goal of the next stage in the design process is to create computational models of behavior from the design specifications created in the previous stage and convert these models into computer algorithms that automatically generate gaze behaviors. The approach to creating these computational representations is to integrate rules from theory (e.g., that a speaker looks at an addressee before a turn as described in the Sacks et al. 1974 article) and findings from empirical data (e.g., that regarding the distribution of the length of a speaker's gaze at an addressee) into hybrid rule-based/ stochastic representations. The three studies presented in this dissertation developed conventional and hierarchical state machines to represent gaze behaviors. However, I envision a machine-learning-based approach to computational modeling and automatic generation of behaviors in the future. This vision is further discussed in Chapter 6.

1.3.3. Experimental Evaluation

The evaluation of the designed behaviors is done through controlled laboratory studies in which human participants are asked to perform an experimental task with robots. These tasks are designed to allow participants to immerse themselves in an experience with a robot. Gill et al. (1998) suggest that people have poor mental representations of those with whom they have little experience; therefore, they cannot confidently make judgments about them. For example, "A hiring committee may be intrigued by a job candidate on skimming [the candidate's] vita but will hire [the candidate] only after boosting its confidence through conversations with the candidate" (Gill et al., 1998). In judging robots, people might rely on similarly poor mental representations because they have very little experience with robots. Therefore, it is assumed here that the more experience people have with a robot, the richer their mental representation of the robot and the more confident and consistent their judgments of the robot would be. Furthermore, Lee et al. (2005) showed that people's beliefs about a humanlike robot's capabilities and knowledge are developed mostly through extrapolating from their own knowledge of people. In this process, they rely on simple cues from the robot such as the robot's origin and language. Therefore, the more cues they get from the robot, the more consistent judgments of the robot would be across individuals.

The experiences with robots are designed to follow social "interaction rituals" (Goffman, 1971) or "social episodes" (Forgas, 1979), particularly rituals of greeting and leave-taking. Goffman (1955) describes greetings as serving "to clarify and fix the roles that participants will take during the occasion of the talk and to commit participants to these roles" and leave-taking rituals as providing "a way of unambiguously terminating the encounter." Following these rituals is particularly important in the context of this research because gaze cues are shown to play a significant role in producing these behaviors (Kendon & Ferber, 1973). Therefore, in their interactions with human participants, robots were modeled to follow greeting and leave-taking rituals.

The technical implementation of the experiments combined automatic algorithms (e.g., gaze generation algorithms) and Wizard-of-Oz techniques (e.g., fixing participants' locations to avoid the cost of implementing face tracking and minimize errors and delays caused by real-time recognition) (Dahlbäck et al., 1993). Implementation details are provided in the study descriptions.

1.4. Contributions

This dissertation has a number of methodological, theoretical, and practical contributions to the design of humanlike robots and research on human communication, particularly to the fields of human-robot interaction (HRI), human-computer interaction (HCI), computer-supported collaborative work (CSCW), and computer-mediated communication (CMC). The methodological contributions include a theoretically and empirically grounded, interdisciplinary, integrated process for designing, building, and evaluating social behavior for humanlike robots. The theoretical contributions advance our understanding of human communicative mechanisms from a computational point of view and of people's responses to theoretically based manipulations in these mechanisms when they are enacted by humanlike robots. The practical contributions include a set of design variables or behavioral parameters—such as gaze target, frequency, and duration—that designers can use to create gaze behaviors for robots that could be manipulated to obtain social and cognitive outcomes and the computational models of social behavior created for the empirical studies. The following sections describe these contributions in detail.

1.4.1. Methodological Contributions

This dissertation makes five chief methodological contributions:

- A methodology for grounding design decisions in theory and empirical data in designing humanlike behavior for humanlike robots (Chapter 1).
- An integrated, interdisciplinary process for designing, building and evaluating communicative mechanisms for humanlike robots (Chapter 1).

- An experimental paradigm for studying how speaker attention can be manipulated through changes in gaze behavior and measuring the effects of different levels of attention on information recall and subjective evaluations of the speaker (Chapter 3).
- An experimental paradigm for studying how speakers can signal conversational roles using gaze cues and for measuring whether people conform to these roles and what effects conforming to these roles have on information recall, task attentiveness, liking, and groupness (Chapter 4).
- An experimental paradigm for studying leakage gaze cues in human communication and human-robot interaction and measuring how the presence and absence of these cues might affect attributions of mental states using task performance measures (Chapter 5).

1.4.2. Theoretical Contributions

The theoretical contributions of this dissertation fall into two categories: (1) contributions to human communication research with new knowledge on human communication mechanisms and (2) those to human-robot interaction with new knowledge on how robots' use of human communication mechanisms leads to social and cognitive outcomes.

1.4.2.1. Human Communication

- An understanding of the spatial and temporal properties of the gaze behavior that accompanies oratorial speech in American English; in particular that speakers in oratorial situations look at the faces of their addressees, spots in the environment, and a fixation point in front of them (Chapter 3).

- An understanding of the spatial and temporal properties of conversational speaker gaze behavior in different role structures in Japanese, particularly gaze targets, frequencies, and fixation lengths of speakers' gaze toward addressees, bystanders, and overhearers (Chapter 4).
- An understanding of speaker gaze cues that help manage turn-exchanges in Japanese (Chapter 4).
- An understanding of speaker gaze cues that signal conversational roles in Japanese, particularly the roles of addressee, bystander, and overhearer (Chapter 4).
- An understanding of speaker gaze patterns that signal the thematic structure of casual conversations in Japanese (Chapter 4).
- An understanding of how speakers leak information about their mental states through gaze cues and attempt to conceal this leakage using gaze cues in the context of a guessing game played in Japanese (Chapter 5).
- Evidence that people use information from others' gaze cues—including leakage gaze cues—to make attributions of mental states (Chapter 5).

1.4.2.2. Human-Robot Interaction

- Evidence that manipulations in robot gaze lead to significant social and cognitive outcomes such as better information recall, stronger conversational participation, and stronger attribution of mental states (Chapters 3, 4, and 5).
- Evidence that how much a robot looks at an individual affects that individual's performance in recalling the information presented by the robot (Chapter 3).

- Evidence of strong gender effects on the perception of gaze behavior, particularly on positive evaluations of the robot (Chapter 3).
- Evidence that turn-taking signals created through robot gaze are correctly interpreted by people and lead to fluid, natural sequences of turn-taking in human-robot conversations (Chapter 4).
- Evidence that gaze cues alone (i.e., whether a robot looks at a person at interactionally significant points in a conversation) can lead to more liking of the robot, higher feelings of groupness with the robot and others in the conversation, and heightened attentiveness to the conversation (Chapter 4).
- Evidence that contextualized leakage gaze cues can communicate mental states of a robot evidenced by improvements in task performance (Chapter 5).
- Evidence that people's interpretations of leakage gaze cues are affected by the physical design of the robot (Chapter 5).
- Evidence that people's interpretations of nonverbal cues in robots are affected by whether they own pets (Chapter 5).
- Evidence that robots can also effectively conceal leaked information, but this behavior negatively affects people's perceptions of the robot, particularly how sociable, cooperative, and helpful they perceive the robot to be (Chapter 5).

1.4.3. Practical Contributions

The following five main practical contributions are made by this dissertation:

- A set of design variables for designing social gaze mechanisms (Chapters 3, 4, and 5).

- A computational model of oratorial speaker gaze behavior that signals information structure represented as a probabilistic state machine and programmed in C++ (Chapter 3.
- A computational model of conversational speaker gaze behavior with gaze mechanisms to help manage turn-exchanges, signal conversational roles, and cue information structure represented as a hierarchical probabilistic state machine and programmed in Java (Chapter 4).
- A computational model of gaze behavior for producing leakage gaze cues and concealing gaze cues at question-answer sequences represented as a probabilistic state machine and programmed in Java (Chapter 5).
- A number of Java-based data analysis tools created for coding and analyzing video, audio, and text data (Chapters 3, 4, and 5).

1.4.4. Design Variables

One of the goals of this dissertation was to extract a number of design variables or behavioral parameters for social gaze that designers can use to create gaze behaviors that could be manipulated to obtain certain social and cognitive outcomes in human-robot interaction. This work identified three sets of design variables for social gaze behavior: temporal variables, spatial variables, and gaze mechanisms. This section provides a brief overview of these sets of variables as a starting point for designing gaze behaviors for robots.

1.4.5. Spatial Variables

The most primitive design variable for social gaze is the gaze target. While the number of gaze targets can be infinite and change based on the frame of reference of the gaze

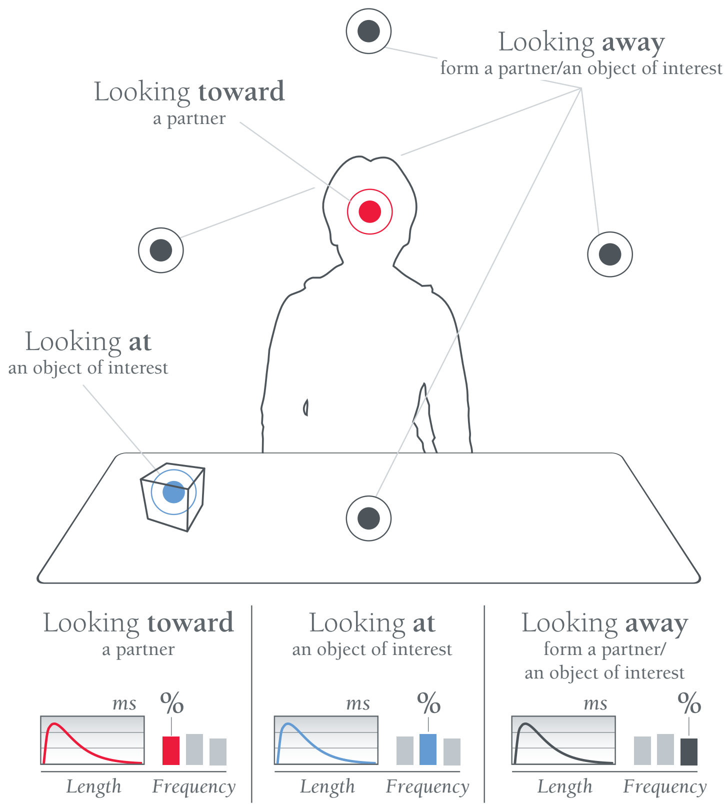


Figure 1.3. An abstract illustration of the spatial and temporal variables of social gaze from the speaker's perspective.

source, I identified three primary areas of gaze targets: toward a partner, at an object of interest, and away from a partner or an object of interest. An abstract illustration of these variables are provided in Figure 1.3.

Looking toward a partner – One of the most salient signals in social interaction is directing gaze at a partner's eye region. Being looked at evokes strong physiological (Coss, 1970) and neurological (Pelphrey et al., 2004) responses and can lead social and cognitive outcomes such as increased attention (Langton et al., 2000) and intimacy (Argyle & Dean, 1965).

Looking at an object of interest – Another important set of gaze targets includes objects of interest or mutual interest in the environment. People tend to follow others' gaze direction because "the direction of a person's gaze usually indicates what object [the person] is interested in or what person [the person] is responding to in the sphere of the environment" (Gibson & Pick, 1963).

Looking away from a partner or an object of interest – Fixation points or random locations in the environment or on a partner's body make up another set of gaze targets. These targets can change based on conversation structure, the content of the conversation, the complexity of the objects of interest in the environment, and the relevance of these objects to the conversation (Argyle & Graham, 1975).

1.4.6. Temporal Variables

Each gaze target shows different characteristics in how frequently and for how long speakers look toward and away from these targets. These differences can be captured with two temporal variables: gaze frequency and length. Figure 1.3 provides an abstract illustration of these variables.

Gaze frequency – How frequently each target is looked at can be defined as a percentage of overall gaze duration during speech and can be used to determine how frequently a robot should look at different targets. Frequencies for each target can be manipulated to orient different levels of attention toward that target.

Gaze length – How long each gaze target is looked at can be captured as a single average duration or as a distribution of gaze durations. These durations can be used to determine how long a robot needs to glance at each target and can be manipulated to change the amount of attention communicated by the robot. A total gaze "amount" for each target can also be parameterized as a function of gaze lengths and frequencies

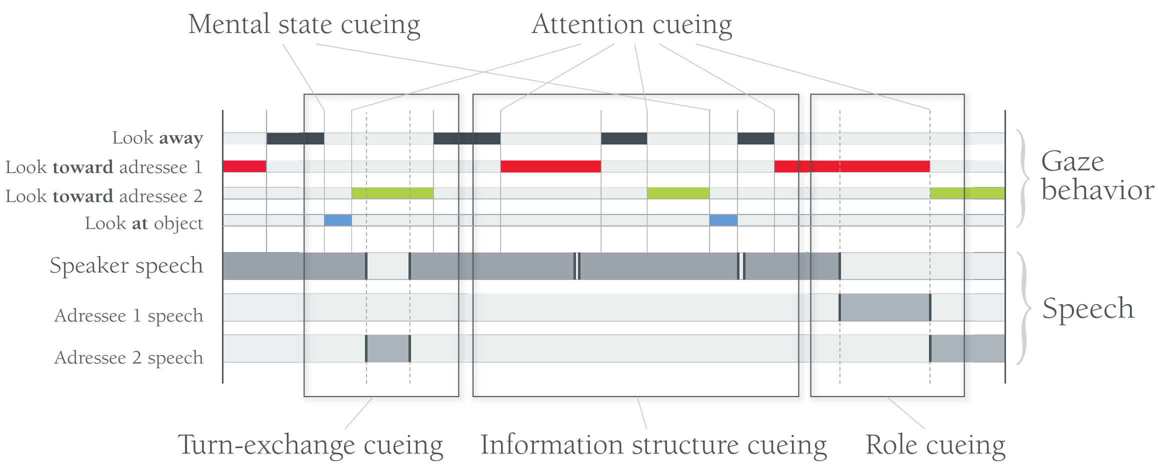


Figure 1.4. An abstract illustration of the gaze mechanisms identified in this dissertation.

and can be useful in manipulating and measuring the amounts of gaze frequencies and gaze lengths for each target.

1.4.7. Gaze Mechanisms

A number of gaze mechanisms embody these spatial and temporal design variables into particular patterns in order to perform communicative functions including cueing attention, signaling information structure, facilitating turn-exchanges, cueing conversational roles, and leaking information on mental states. Figure 1.4 provides an abstract illustration of these mechanisms mapped on a timeline.

Gaze cueing of attention – Looking toward a gaze target communicates attending to that target and, when performed in particular patterns and at interactionally significant points in a conversation, can communicate information about discourse, conversation, and role structures and complex mental states such as emotions, beliefs, and desires. For instance, looking toward one addressee among multiple addressees can indicate a primary communication target and place others at a secondary stature. This might also affect how much attention the speaker gets from each addressee.

Gaze cueing of information structure – Speakers shift their gaze in particular patterns based on the structure of their speech. For instance, the majority of utterances in English start with looking away from an addressee, followed by a gaze shift toward the addressee at a structurally significant point of the utterance.

Gaze cueing of turn-exchanges – Gaze shifts can also help speakers coordinate turn-exchanges using a sequence of three signals: a turn-yielding signal at the end of a turn, a turn-taking signal while waiting to take the floor, and another floor-holding signal after taking the turn.

Gaze cueing of conversational roles – Gaze cues can also be used in cueing the roles of conversational partners. These cues are particularly suggestive of conversational roles when they are produced at interactionally significant points of the conversation. For instance, directing gaze toward one conversational partner and away from another partner at the transition from greetings to the body of the conversation might signal the former the role of addressee and the latter the role of bystander.

Gaze cueing of mental states – Gaze cues, when combined with contextual information, can signal mental states. For instance, glancing at one object among other objects might communicate a preference or emotion toward that object.

This chapter described the scenarios that outline the research context and motivate research questions. An overview of the approach taken to addressing these questions and a summary of contributions are also provided. The next chapter provides a review of related work on gaze from research on human communication, humanlike virtual agents, and humanlike robots focusing specifically on the social contexts outlined by the scenarios presented in this chapter.

2. Background

Research in human-computer interaction has shown that people respond to computers in social ways (Nass & Steuer, 1993; Reeves & Nass, 1996; Sproull et al., 1996). Nass et al. (1993) proposed a framework called "Computers as Social Actors" (CASA) and showed in a series of studies that people make attributions of gender stereotypes (Nass et al., 1997) and personality to (Moon & Nass, 1996), respond to flattery (Fogg & Nass, 1997) and humor from (Morkes et al., 1999), give credit to (Moon & Nass, 1998), and show politeness towards (Nass et al., 1999) computers. They argued that these responses are mindless and automatic (Nass & Moon, 2000) following the proposition that people automatically respond to relevant social stimuli (i.e., when the stimulus follows common norms and patterns of interaction) (Langer et al., 1978; Bargh et al., 1996).

Cassell (2001) argues that humanlike stimuli are more likely to evoke social responses than machinelike stimuli, because people have a propensity to seek an embodiment for intelligence and a social locus of attention. In support of this argument, Sproull et al. (1996) showed that explicit humanlike cues such as a humanlike face presented on a computer screen as compared with a text-based computer led people to make stronger attributions of personality to the computer and present themselves more positively to it, and feel more relaxed and assured by the computer. These results suggest that humanlike cues provide a sense of presence and disambiguate what communicative channels are open to people (e.g., speech, gaze, facial expressions, gesture, etc.), making communication more fluent and allowing people to have a more certain mental representation of the computer (Kiesler, 2005).

This dissertation builds on this research, but focuses particularly on understanding how people respond to humanlike gaze. It is informed by literature on social gaze behavior in human communication research, embodied conversational agents, and robotics. The literature survey below encompasses a review of related work from all three literatures, with a focus on the social contexts provided by the scenarios presented in the previous section.

2.1. Gaze Cues in Social Interaction

During social interaction, people look at others for an average of 61% of the time—longer than they speak (Argyle & Ingham, 1972). Through gazing at others, people study others' behavior and appearance and look particularly in the region of their eyes (Cook, 1977). The eyes are such an important source of social information that even infants aged four weeks are able to locate the eyes of an observer (Wolff, 1963). Newborns prefer faces with visible eyes (Batki et al., 2000) and moving pupils (Farroni et al., 2002). For any social interaction to be initiated and maintained, parties need to establish eye contact. Goffman (1963) argued that, through establishing eye contact, people form "an ecological eye-to-eye huddle" through which they signal each other that they agree to engage in social interaction. Simmel (1921), as quoted in (Argyle & Cook, 1976), describes this mutual behavior as "a wholly new and unique union between two people [that] represents the most perfect reciprocity in the entire field of human relationship."

Research has shown that people are extremely sensitive to being looked at (Gibson & Pick, 1963). This sensitivity may have evolved as a survival mechanism to detect whether a predator is attending (Emery, 2000). Neurophysiological evidence shows that this mechanism might be supported by a dedicated 'eye direction detector' in the brain (Baron-Cohen, 1995). Pictures of eyes (Bateson et al., 2006)—even simulated "eyespots" on a computer screen (Haley & Fessler, 2005)—are found to influence people's decision-making behavior. On the road, drivers and pedestrians move off more rapidly from stoplights when they are stared at (Ellsworth et al., 1972). Drivers are more likely to stop for hitchhikers who establish eye contact with them (Snyder et al., 1974).

Research on gaze, which started in early 1960s, has shown that gaze behavior is tightly intertwined with many other aspects of social interaction. Gender (Exline, 1963; Argyle & Ingham, 1972; Bayliss et al., 2005), personality differences (Kleck & Nuessle, 1968; Mobbs, 1968; Strongman & Champness, 1968), conversational role (Exline & Winters, 1966; Kendon, 1967), the topic of conversation (Exline, 1963; Exline et al., 1965a; Abele, 1986), whether interaction takes place in public (Goffman, 1963; Kendon, 1973; Kendon & Ferber, 1973), the familiarity of the parties (Exline, 1963; Noller, 1984), and many other factors are found to affect gaze behavior. The tight coupling between gaze behavior and many other aspects of social interaction has made the study of gaze behavior central to social psychology. Argyle and Cook (1976) argue, "Any account of social behavior which fails to deal with the phenomena of gaze is quite inadequate."

2.1.1. What is Gaze?

Most human communication research literature on gaze is concerned with the direction of eyes. While eyes are the primary source of information about the direction of attention, social gaze involves a complex coordination of the eyes, the head, and body orientation that is sensitive to the social context (Emery, 2000; Frischen et al., 2007). For instance, when cues from eyes and head are congruent, people can interpret direction of attention faster than they can when cues are incongruent (Langton & Bruce, 1999).

Without information from the eyes, cues such as head orientation, body posture, and pointing gestures might also indicate direction of attention (Langton et al., 2000). Research on neurophysiological aspects of gaze has shown that signals created in part of the brain that is responsible for processing social information by observing eye direction are stronger than those evoked by observing head direction, which suggests that directional information from gaze, head, and body cues might be combined hierarchically in a mechanism dedicated to detect another's direction of attention (Perrett et al., 1992). The existence of this hierarchy is supported by behavioral evidence. When people are at greater distances, head orientation becomes a stronger cue than information from the eyes in determining direction of attention (Von Cranach & Ellgring, 1973).

In addition to this hierarchical relationship, eye and head orientation might convey different social information. For instance, Gibson and Pick (1963) showed that participants misjudge the gaze direction of a confederate when the confederate's head is not facing them. Hietanen (2002) argues that this misjudgment is because "facing away" might be interpreted as "socially disinterested" delaying the processing of eye direction, supporting the hypothesis that head orientation and eye direction convey different messages.

2.1.2. Definitions of Social Gaze Behavior

A review of literature on gaze from different perspectives reveals a number of terms, concepts, and social situations where gaze plays a significant role. Below, definitions are provided for concepts that are relevant to this thesis.

Gaze, **one-sided gaze**, **eye-gaze**, **looking at**, **visual orientation towards –** A looks at B in or between the eyes, or, more generally, in the upper half of the face (Cook, 1977).

Mutual gaze, **eye contact –** Both A and B look into each other's face, or eye region, thus acting simultaneously as sender and recipient (Von Cranach & Ellgring, 1973).

Averted gaze, **gaze avoidance**, **gaze aversion**, **cutoff –** A avoids looking at B especially if being looked at, and/or moves the gaze away from B (Von Cranach & Ellgring, 1973; Emery, 2000).

Gaze following – A detects B's direction of gaze and follows the line of sight of B to a point in space (Emery, 2000).

Joint attention, **visual co-attention**, **deictic gaze –** A follows B's direction of attention to look at a fixed point in space (such as an object) (Butterworth, 1991).

Shared attention – Both A and B look at a fixed point in space and are aware of each other's direction of attention (Baron-Cohen, 1995; Emery, 2000).

2.1.3. Gaze and Speech

Research on conversational functions of gaze show that gaze behavior is closely linked with speech (Argyle & Cook, 1976). Kendon (1967) identified patterns in speakers' and addressees' gaze during role exchanges. For instance, he found that speakers mostly look away from their addressees at the beginning of an utterance, but look at their addressees at the end of an utterance. As parties switch roles at the beginning of the next utterance, the new speaker looks away from the new addressees (Kendon, 1967). In this situation, looking away at the beginning of an utterance and during hesitant speech indicates holding the floor (Nielsen, 1962; Kendon, 1967) and serves to avoid information overload in the planning of the utterance (Goodwin, 1981). Looking at the addressee at the end of an utterance, on the other hand, communicates that the speaker is ready to pass the floor to the addressee (Nielsen, 1962).

The information structure of a speaker's utterances is also account for gaze shifts (Cassell et al., 1999b). When utterances are looked at as theme-rheme progressions (Halliday, 1967), at the beginning of each theme, speakers look away from the addressees 70% of the time, and at the theme-rheme junction, they look at their addressees 73% of the time (Cassell et al., 1999b).

The remainder of this review focuses on three social functions of gaze behavior following the three scenarios presented earlier; communication of attention, regulating conversational participation, and communicating mental states. Related work on each function involves reviews of relevant literature on human communication, humanlike virtual agents, and humanlike robots.

2.2. Gaze Cues in the Communication of Attention

An important aspect of human cognitive system is the ability to orient attention to information in the environment that is relevant to one's behavioral goals (Posner, 1980; Frischen et al., 2007). One of the most salient cues of this orientation is gaze direction (Baron-Cohen, 1995; Emery, 2000). People direct their gaze at each other to signal that their attention is directed at the other (Goffman, 1963). Being looked at by another produces an immediate heightening of arousal (Nichols & Champness, 1971; Patterson, 1976; Kleinke, 1986). Neurophysiological research suggests that this response is induced by an “eye-direction detector” (Baron-Cohen, 1995) or a “direction of attention director” (Perrett & Emery, 1994) located in the Superior Temporal Sulcus (STS), the part of human brain involved in deriving social meaning (Perrett et al., 1992; Baron-Cohen, 1995). The STS responds to being looked at more than it does to other gaze stimuli, as it is engaged in the processing of the social information (Pelphrey et al., 2004). This processing funnels the attention to the looker and might delay any response to other stimuli (Senju & Hasegawa, 2005).

2.2.1. Attention and Learning

The establishment of mutual orientation of attention can lead to significant social outcomes such as increased intimacy (Patterson, 1976), attraction (Exline & Winters, 1966; Mason et al., 2005; Frischen et al., 2007), and attention (Langton et al., 2000), all of which contribute to increased attention on and better recall of verbal communication (Exline & Eldridge, 1967; Fry & Smith, 1975; Otteson & Otteson, 1980; Sherwood, 1987). For instance, Otteson and Otteson (1980) conducted a study in which they asked a teacher to tell stories to 46 primary school students in groups of four, manipulated the teacher’s gaze to be directed toward only two of the students, and found that those toward whom the teacher gazed had better recall of the story than others did. Sherwood (1987) replicated these results with college students. Fry and Smith (1975) showed that participants performed better in a digit-encoding task when the instructors made as much eye contact as possible while reading the instructions than when they made as little eye contact as possible.

In the classroom, significant differences were observed between experienced and/or effective teachers and inexperienced and/or ineffective teachers in the frequency of direct eye contact with students during the first week of class (Brooks, 1985). Woolfolk and Brooks (1985) suggest that teachers consciously and explicitly use gaze cues to attract the attention of their students. In fact, 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).

2.2.2. Person Perception

Gaze cues are also used in evaluations of personality (Goffman, 1963; Kleck & Nuessle, 1968; Kendon & Cook, 1969; Cook & Smith, 1975). In general, positive evaluations of a partner increase consistently with the amount of gaze from zero to normal but decrease with too much gaze (Argyle et al., 1971). People who look at others 80% of the time are rated as more friendly, self-confident, natural, mature, and sincere, while those who look at others 15% of the time are perceived as cold, pessimistic, cautious, nervous, defensive, immature, evasive, submissive, indifferent, sensitive, and lacking confidence (Kleck & Nuessle, 1968; Cook & Smith, 1975).

People are sensitive to not only the amount of gaze they receive but also the patterns of gaze. We expect others to look in certain ways and are disturbed when we encounter unusual gaze patterns (Goffman, 1963). People who look in long, infrequent gazes are preferred over those who look in short, frequent ones (Kendon & Cook, 1969). People who are observed to move their eyes to establish eye contact are evaluated more likable than those who are observed to break eye contact (Mason et al., 2005).

The feeling of being looked at is also shown to affect economic decisions (Haley & Fessler, 2005; Bateson et al., 2006; Burnham & Hare, 2007). For instance, Bateson and others (2006) placed images of either a pair of eyes or flowers on an “honesty box” that was used at a school cafeteria to collect money for drinks and found that people paid nearly three times as much for their drinks when they saw images of eyes than when they saw images of flowers. Similar results were found in a study where participants saw images of MIT’s Kismet robot on their computer screen (Burnham & Hare, 2007) and in another study where their computer backdrop contained schematic eyes (Haley & Fessler, 2005).

2.2.3. Gaze Behavior as an Intimacy-Regulation Mechanism

While increased gaze can effectively evoke a number of social and cognitive outcomes, too much gaze might lead to discomfort in interaction partners due to increased intimacy and prompt them to compensate for this increase (Argyle & Dean, 1965). The “Intimacy Equilibrium Theory” suggests that interaction partners develop equilibrium for interpersonal intimacy, which is a function of a number of communicative mechanisms including gaze, physical proximity, intimacy of the conversational content, and amount of smiling (Argyle & Dean, 1965). For any set of participants, the level of intimacy is at a certain degree and participants try to keep this degree constant over the course of the interaction. When one of the components in the model changes (e.g., an increase in physical proximity), people tend to maintain the equilibrium by shifting one or more of the other components in the reverse direction.

Argyle and Dean (1965) experimentally demonstrated that people reduce how much they look toward their partners at closer distances. Kendon (1967) found that smiling and the amount of gaze were inversely correlated, confirming Argyle and Dean's theory. Exline and others (1965b) showed that intimacy of the conversational content affected the amount of gaze. Their results were partly consistent with Argyle and Dean's theory; participants looked at their partners more when they were speaking, but not while listening. One explanation of this effect is that speaking about intimate topics evokes more embarrassment than listening to others talk about intimate topics (Argyle & Cook, 1976). On the other hand, Abele (1986) manipulated the amount of intimacy in a conversation and found that this did not affect how much participants looked at their partner. Although she found that when partners talked about intimate topics, the level of gaze increased over the course of the conversation, while it decreased when participants talked about non-intimate topics. These results suggest that people do regulate their intimacy with their partners when they talk about intimate topics. However, people might also adapt to the changes in the level of intimacy during a conversation.

2.2.4. Gaze Cues and Communication of Attention in Humanlike Virtual Agents

A number of studies looked at communication of attention in the design of humanlike virtual agents (Khullar & Badler, 2001; Peters & O'Sullivan, 2003; Peters, 2005). However, these studies focused mostly on the automatic production of gaze cues to communicate an agent's direction of attention based on the detection of salient stimuli in the environment or the attention level of a conversational partner, and did not consider how different aspects of the agent's gaze affected the partner's attention level or perceptions of and interactions with the agent. For instance, Khullar and Badler (2001) developed a model for automatically animating an agent's gaze direction based on the outputs of different cognitive mechanisms such as visual search and tracking. Peters (2005) developed a virtual agent with the ability to interpret its partner's level of interest and use gaze cues to provide feedback on its attention to maintain conversational flow.

In one study, Bailenson et al. (2005) looked at how gaze cues of a speaker could be "augmented" in an immersive virtual environment to direct the attention of the speaker towards two listeners simultaneously. They compared participants' evaluations of the speaker across augmented and normal gaze conditions and found that women agreed with the speaker's message more in the augmented gaze condition than in the normal gaze condition.

2.2.5. Gaze Cues and Communication of Attention in Humanlike Robots

Two studies in human-robot interaction looked at communication of attention. Bruce et al. (2002) designed a social robot called Vikia with the ability to detect and orient its gaze at passersby in a hallway. They evaluated whether the ability to orient attention towards a person would increase people's likelihood to interact with the robot and found that passersby were more likely to stop when the robot oriented its attention at people than when it did not. Imai et al. (2002) conducted an experiment where they seated eight participants in a circle and placed a humanlike robot at the center of the circle. As the robot oriented its gaze direction at different people, participants were asked whom they thought the robot was attending to. Their results showed that people had a high sensitivity to being looked at and could identify when the robot looked at them with 91% accuracy and when it looked at the person sitting next to them with 80% accuracy. These two studies provide evidence that gaze cues can communicate a robot's direction of attention and people respond to them as valid stimuli.

2.2.6. Summary

The literature reviewed in this section highlights the central role that gaze cues play in communicating attention and the strong propensity humans have to automatically detect gaze and infer one's direction of attention. Detecting that another individual is directing attention toward us can evoke particular outcomes including better learning of the information presented by the individual, more positive evaluations of this person, and more generous economic decisions. Furthermore, effective communicators consciously and explicitly manipulate their gaze behaviors to obtain these outcomes, such as a teacher building more eye contact to improve student learning. On the other hand, too much gaze might evoke discomfort and prompt gaze targets to compensate for the increased intimacy.

This review also suggests that humanlike virtual characters in immersive virtual environments and humanlike robots can use gaze cues to communicate their direction of attention and that virtual agents can evoke social responses through increased gaze such as more agreement in the agent's messages. However, whether manipulations in gaze cues might lead to social and cognitive outcomes such as better learning or more positive evaluations with robots has not been studied.

The next section provides an overview of literature on the *conversational* use of gaze cues, particularly in establishing and negotiating participant roles in human conversations and conversations with humanlike virtual agents and humanlike robots.

2.3. Gaze Cues in Signaling Conversational Participation

In conversations, people work together as participants (Clark, 1996). The roles of the participants, a phenomenon described by Goffman (1979) as "footing," and how these roles might shift during social interaction, are particularly important in understanding spoken discourse (Hymes, 1972; Hanks, 1996). At the core of these roles are those of the speaker and the addressee (Clark, 1996). While these roles might be fixed in some social settings (e.g., lectures), most conversational settings allow for shifting of roles. At any "moment" (Goffman, 1979) in a two-party conversation, one of the participants plays the role of the speaker and the other plays the addressee. Conversations with more than two participants also involve side participants who are the "unaddressed recipients" of the speech at that moment (Goffman, 1979; Wilkes-Gibbs & Clark, 1992; Clark, 1996).

In addition to these "ratified participants" (Goffman, 1979), conversations might involve "non-participants" (Clark, 1996). For instance, there might be bystanders whose presence the participants acknowledge and who observe the conversation without being participants in it (Goffman, 1979; Clark & Carlson, 1982; Clark, 1996). There might also be hearers whose presence the participants do not acknowledge, but who follow the conversation closely, such as overhearers who are unintentionally listening to the conversation and eavesdroppers who have engineered the situation to purposefully listen to the conversation (Goffman, 1979). Figure 2.1 provides an abstract illustration of these different levels of participation.

The direction of gaze plays an important role in establishing and maintaining conversational participant roles. In conversations that involve more than two people, the gaze of a speaker towards another participant can signal that the speaker is addressing that participant (Sacks et al., 1974; Goodwin, 1981). In this situation, the speaker indicates a "communication target" (Bales et al., 1951). When there is no intended target (i.e., when a speaker is addressing a group), gazing at a participant long enough might create the belief that the speaker is addressing that participant (Bales, 1970). On the other hand, when there is an intended target and the speaker does not signal by means of gaze who is being addressed, breakdowns might occur in the organization of the conversation (Schegloff, 1968). Furthermore, apparently intentional avoidance of gaze in situations where acknowledgements are expected can lead to feelings of ostracism (Williams, 2001).

Gaze direction also serves as an important cue in shifting roles during turn-exchanges (Nielsen, 1962; Duncan, 1974; Sacks et al., 1974; Goodwin, 1980; Goodwin, 1981) and overlapping talk (Schegloff, 2000). For instance, speakers might look away from

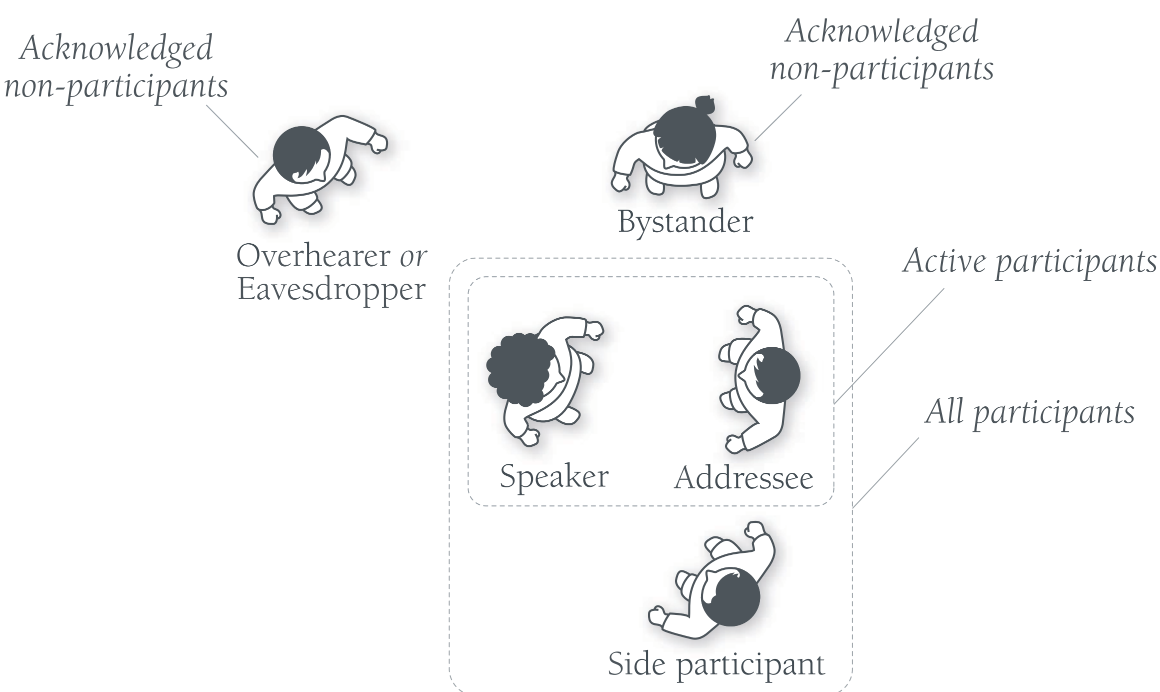


Figure 2.1. Levels of conversational participation (adapted from Goffman, 1979; Clark, 1996).

their addressees to indicate that they are in the process of constructing their speech and do not want to be interrupted and look at their addressees to signal the end of a remark and the passing of the floor to another participant (Nielsen, 1962). In this context, the participant at whom a speaker looks at the end of a remark would be more likely to take the role of the speaker next (Weisbrod, 1965 as described in Kendon, 1967). Shifting of roles might be delayed when remarks do not end with gazing at another participant (Kendon, 1967; Vertegaal et al., 2000). When gaze levels are particularly low, such as in a conversation between strangers, gaze plays an especially important role in cueing role exchanges (Beattie, 1980).

2.3.1. Gaze Cues and Conversations with Humanlike Virtual Agents

In human-computer interaction research, conversational gaze cues have been extensively studied in the context of designing embodied conversational agents (Cassell et al., 1994; Cassell et al., 1999a; Garau et al., 2001; Thorisson, 2002; Heylen et al., 2005; Rehm & Andre, 2005). Cassell et al. developed a number of systems that use verbal and nonverbal behaviors to support conversational mechanisms such as turn-taking, feedback, repair, synchronized speech, and intonation (Cassell et al., 1994; Vilhjalmsson & Cassell, 1998; Cassell et al., 1999a). While these systems combined nonverbal cues such as gaze, facial expressions, hand gestures, and postural shifts in the design of the agent (Cassell et al., 1999a; Cassell, 2001), gaze cues were considered as the most salient signal to establish conversational roles and regulate turn-taking (Cassell et al., 1999b; Vertegaal et al., 2001). Furthermore, research in this area has shown that signals that are designed to resemble human gaze behavior (as opposed to randomly generated signals) lead to more efficient conversations, better task performance, and more positive evaluations of the agent (Colburn et al., 2000; Garau et al., 2001; Heylen et al., 2005).

Another area in which conversational gaze cues have been studied is research in immersive virtual environments. Two studies in this area focused on understanding how these cues might shape participant roles and to how different forms of participation might affect the social outcome of human-agent conversations (Bailenson et al., 2005; Rehm & Andre, 2005). As described in the previous section, Bailenson et al. (2005) studied how speaker gaze cues might be "augmented" to signal to two speakers simultaneously that they are the agent's only addressee. They found that the impression of being the primary addressee of a speaker led the participants to agree more with the speaker's messages. Rehm and Andre (2005) asked two participants to play a game with a virtual character in which each player played the roles of the speaker and the addressee and evaluated people's involvement in the conversation. Their results suggest that, when appropriate cues are present, people conform to the participant roles that an agent signals to them.

2.3.2. Gaze Cues and Conversations with Humanlike Robots

In research in human-robot interaction, a more recent but growing body of literature focuses on conversational use of gaze cues (Matsusaka et al., 2001; Bennewitz et al., 2005; Kuno et al., 2007; Trafton et al., 2008; Yamazaki et al., 2008). Among these, a few studies have examined the conversational effectiveness of robot gaze, particularly in regulating turn-taking in two-party (Kuno et al., 2007; Yamazaki et al., 2008) and multi-party conversations (Matsusaka et al., 2001; Bennewitz et al., 2005; Trafton et al., 2008). Kuno and others (2007) developed gaze behaviors for a museum guide robot that looked at its addressee at "turn-relevant places" (Sacks et al., 1974) to regulate turn-taking. Yamazaki and others (2008) showed in an experiment that looking at an addressee at turn-relevant places evoked more backchannel responses in the addressee than looking at random places. Matsusaka et al. (2001) and Bennewitz et al. (2005) developed robots that could participate in multi-party conversations following the turn-taking model suggested by Sacks et al. (1974) for human conversations. Trafton et al. (2008) developed appropriate listening behaviors for a robot as a bystander and experimentally showed that people found the robot's gaze behavior to be more natural when the robot looked at the speaker only during turns as opposed to during turns and backchannel responses.

2.3.3. Summary

The literature reviewed in this section illustrates the central place that footing has in conversations and the essential part that gaze cues play in signaling footing. These cues provide information about who the acknowledged participants of a conversation are, who is being addressed, and who should take the next speaking turn. When these cues are absent, breakdowns occur in the organization of the conversation. Furthermore, intentional avoidance of the production of these cues can engender outcomes such as ostracism.

Research on humanlike virtual agents suggest that agent gaze cues can lead to different forms of participation and social outcomes such as better task performance, increased agreement with the agent, and more positive evaluations of it. Human-robot interaction studies also provide strong evidence that robot gaze might support core conversational functions such as turn-taking. However, whether a robot's gaze cues can shape conversational participation or lead to different social outcomes remains unexplored. Furthermore, whether results obtained with humanlike virtual agents might transfer to interactions with humanlike robots is unknown.

In the next section, literature on the use of gaze cues in communicating mental states is reviewed. A review of related work on the design of humanlike virtual agents and humanlike robots is also provided.

2.4. Gaze Cues in Attributions of Mental States

Human communication involves a number of nonverbal cues that are seemingly unintentional, unconscious, and automatic—both in their production and perception—and convey rich information about the mental and emotional states of communication partners. A particular family of such cues is "nonverbal leakage," as termed by Ekman and Freisen (1969), in which individuals "leak" their feelings or thoughts through nonverbal cues that they unintentionally produce due to heightened arousal, feelings such as guilt, attempts to control their behaviors and feelings, and the cognitive complexity of their task such as manufacturing a lie (Zuckerman et al., 1981; Buller et al., 1996; DePaulo et al., 2003).

Research in this area has found that a number of intentional and affective states can be identified simply through observations of leakage cues. For instance, cues from the face, arms, and legs were found to reveal deception and self-deception (Ekman & Friesen, 1969). Furthermore, research suggests that people show an automatic and unconscious propensity to search for cues that might leak information in others' nonverbal behaviors and respond to them (Surakka & Hietanen, 1998; Williams et al., 2001). For instance, studies of smiling showed that people automatically fixate and read cues from the region of the eyes, particularly the "crow's feet" area (Williams et al., 2001), to distinguish genuine smiles—called the "Duchenne Smile" (Ekman et al., 1990)—from smiles of appeasement (Surakka & Hietanen, 1998). Furthermore, naïve observers with no particular expertise can identify these cues and interpret their meanings. For instance, clinical psychological research has shown that using nonverbal leakage cues alone—particularly those from the hands, eyes, mouth, and torso—naïve observers are able to identify the presence and discriminate among varying intensities of anxiety (Waxer, 1977). Similarly, Feldman et al. (1978) showed that naïve observers could distinguish genuine or dissembled praise based on the amount of smiling, instances of pauses in speech, and mouth expressions of the person providing the praise. Chawla and Krauss (1994) found that naïve observers were able to distinguish rehearsed speech from spontaneous speech with reliably higher accuracy than chance using only nonverbal cues. Finally, naïve participants who were asked to review videotapes of a performer reading friendly, neutral, and hostile messages in a friendly, neutral, and hostile nonverbal style were found to rely on nonverbal cues significantly more than the verbal content in their ratings of the messages (Argyle et al., 1971).

2.4.1. Gaze Cues as a Channel of Nonverbal Leakage

Gaze cues are a particularly important set of leakage cues that provide a wealth of information about the mental and emotional states of an individual (Hemsley & Doob, 1978; Kraut & Poe, 1980; Kleinke, 1986; Perrett & Emery, 1994; Baron-Cohen, 1995; Emery, 2000; Blakemore & Decety, 2001; Freire et al., 2004; Frischen et al., 2007). Social and developmental psychological studies have shown that through observing others' gaze patterns, people infer personality traits (Kleinke, 1986)—particularly trustworthiness (Bayliss & Tipper, 2006)—and detect and infer deception (Hemsley &

Doob, 1978; Kraut & Poe, 1980; Leslie, 1987; Baron-Cohen, 1995). For instance, Freire et al. (2004) showed that children as young as four years old could locate a hidden object, using only gaze cues of a performer, despite that they were given verbal information that contradicts the information from the gaze cues.

Neurophysiological research further explains human sensitivity to gaze cues and the automatic propensity to attribute mental states based on information from these cues (Perrett & Emery, 1994; Baron-Cohen, 1995; Emery, 2000; Blakemore & Decety, 2001). Emery (2000) suggests that people combine information from gaze cues with "higher-order cognitive strategies (including experience and empathy) to determine that an individual is attending to a particular stimulus because they intend to do something with the object, or believe something about the object"—an ability called "mental state attribution" or "theory of mind." Baron-Cohen (1995) proposed that the ability to use gaze information to attribute mental states is supported by the interaction between dedicated brain mechanisms such as an "eye-direction detector" and "intentionality detector." Later studies provided support for this proposal by showing that perception of gaze direction activates the same areas of the brain that are involved in making attributions of intention and beliefs (Calder et al., 2002; Castelli et al., 2002). Similarly, research has also found behavioral evidence that people's motor intentions can be inferred by monitoring their gaze direction (Castiello, 2003; Pierno et al., 2006).

2.4.2. Leakage Gaze Cues in Humanlike Virtual Agents

Few researchers have looked at whether mental and emotional states in humanlike virtual agents could be communicated through implicit, seemingly unintentional cues (Kenny et al., 2007; Bailenson et al., 2008; Kenny et al., 2008). Bailenson et al. (2008)

asked participants to interact with an agent that mimicked participants' nonverbal behavior—a common unconscious behavior seen in human communication called the "chameleon effect" (Chartrand & Bargh, 1999)—and to rate whether they thought that the agent was a human or a computer (as in a "Turing Test"). They found that participants rated the agent as human more when the agent mimicked their nonverbal behavior than when it did not do so, suggesting that seemingly unintentional cues affect people's social judgments of virtual agents. Kenny et al. (2007) have developed a "virtual patient" for clinical therapist training with the ability to produce the symptomatic verbal and nonverbal behaviors including gaze of patients with mental disorders such as conduct disorder and post-traumatic stress disorder (PTSD). The evaluation of their system showed that these behaviors induced conversations between the patient and therapist on the main categories of topics associated with these disorders such as "avoidance" and "re-experiencing traumatic events" (Kenny et al., 2008).

2.4.3. Leakage Gaze Cues in Humanlike Robots

Research in human-robot interaction has focused mainly on creating explicit expressions of mental and emotional states (Breazeal & Scassellati, 1999; Scheeff et al., 2000; Breazeal, 2001; Bartneck et al., 2004; Miwa et al., 2004; Gockley et al., 2006) and has not looked at whether these states could be communicated through implicit, seemingly unintentional cues.

2.4.4. Humanlikeness and Perceptions of Behavioral Cues

The physical and behavioral characteristics of a robot, particularly its humanlikeness, might affect how people read and interpret nonverbal cues in robots. Research in virtual agents has shown that the humanlikeness of an agent affects people's social judgments of the agent (Nowak, 2001; Nowak & Biocca, 2003). People reliably rate agents with highly humanlike features to be more socially attractive, more satisfactory as partners (Nowak, 2001), more co-present (Nowak & Biocca, 2003), and more likeable (Parise et al., 1998) than agents with less humanlikeness. They also cooperate more with humanlike agents (Parise et al., 1998).

Research in human-robot interaction has shown similar attributions to robots (Kiesler & Goetz, 2002; Goetz et al., 2003; Hinds et al., 2004). Kiesler and Goetz (2002) argued that humanlike characteristics might engender a more human mental model of a robot. Hinds et al. (Hinds et al., 2004) showed that people took less personal responsibility in a task in which they collaborated with a humanlike robot than in a task in which they collaborated with a machinelike robot, suggesting that people might associate humanlikeness with more competence. Goetz et al. (2003) found that people expected the behavior of a robot to match its task. They complied more with instructions given by a robot that met their expectations of appropriateness than one that did not. This research suggests that people might expect and correctly interpret leakage cues when robots with the appropriate level of humanlikeness enact them.

2.4.5. Summary

This section reviewed research on a behavioral process, called "nonverbal leakage," in which individuals unintentionally "leak" information through nonverbal cues about their emotional and mental states. This literature suggests that observers automatically seek and interpret these cues to make inferences of cognitive processes such as deception. Gaze provides a particularly salient source for nonverbal leakage and elicits an automatic propensity to attribute mental states.

Research in humanlike virtual agents suggests that agents' use of leakage cues can lead to social outcomes such as higher ratings of the agent's humanness. However, gaze cues as source for nonverbal leakage has not been studied in the context of humanlike virtual agents. Research in human-robot interaction has not looked at leakage cues. However, research in this area predicts that the humanlikeness of the robot might affect how people interpret these cues.

Research on gaze suggests that gaze behavior differs greatly across individuals and populations. The next section outlines these differences with particular focus on cultural and gender-based differences.

2.5. Gaze Cues and Interpersonal Differences

Gaze behavior is found to be extremely sensitive to individual differences (Argyle & Cook, 1976), particularly gender (Hall, 1984) and cultural differences (Watson, 1970). Below, a review of these differences is provided.

2.5.1. Gender Differences

In all measures of gaze, women are found to look more than men (Argyle & Cook, 1976; Francis, 1979; Hall, 1984). Argyle and Ingham (1972) found that parties in female dyads looked more (66% vs. 56%) and longer (3.12 vs. 2.45 seconds on average) at each other than parties in male dyads. They also established more (38% vs. 23%) and longer (1.42 vs. 0.86 seconds an average) mutual gaze than parties in male dyads. Similarly, parties in female triads were found to look more (37% vs. 23%) and establish more mutual gaze (8% vs. 3%) than parties in male triads do (Exline, 1963). In unfocused interactions, females are looked at more than males are (Coutts &

Schneider, 1975). Patterson et al. (2002) found passersby in public look more at female confederates (47%) than at male confederates (30%).

Gender-based differences in gaze behavior appear as early as infancy. Girls at 12 and 24 months build more eye contact with their mothers than boys at that age do (Lutchmaya et al., 2002). Similarly, 12-month-old girls show stronger joint attention abilities than boys at that age do (Olafsen et al., 2006). Furthermore, Kagan and Lewis (1965) found that, at six months, girls attend to faces more than boys do. At birth, boys look longer than girls at a mobile, while girls at this age look longer than boys at a face (Connellan et al., 2000). Similarly, Lutchmaya and Baron-Cohen (2002) showed 12-month-old infants videos of moving cars and faces and found that attention in males was more drawn by moving cars and in females by moving faces. All three studies suggest that men and women respond differently to social and non-social stimuli.

2.5.2. Cultural Differences

People from certain cultures are found to look more than others. Watson (1970) found differences in how much people from “contact” cultures (i.e., Arabs, Latin Americans, and Southern Europeans) and those from “non-contact” cultures (i.e., Asians, Indians-Pakistanis, Northern Europeans) look at their partners and found that the former looked at others more than the latter. Japanese were found to look more frequently but with shorter glances than Australians (Elzinga, 1978). Ingham (1972, as described in Argyle and Cook, 1976) compared Swedes and Englishmen and found that the former looked less often (8 vs. 13 glances per minute on average) but with longer glances (5 vs. 2.93 seconds on average).

2.5.3. Individual Differences

While sex and cultural differences account for some of the variability in gaze behavior, greater differences are observed among individuals (Nielsen, 1962; Kendon, 1967; Ellsworth & Ludwig, 1972). Nielsen (1962) found that the amount of time spent looking at a confederate varied between 8% and 73% with an average of 50%. Argyle and Cook (1976) argue that such large variability might result from differences in need for affiliation (which accounts for some of the sex differences), personality type, and varying levels of tolerance for arousal and intimacy.

2.5.4. Summary

The research reviewed in this section highlights the sensitivity of gaze behavior across individuals and populations, particularly the effects of gender and culture in gaze production. Overall, women look more at others than men do in social encounters. Gender configuration also affects gaze production; members of all-female dyads and triads look more at each other than members of all-male dyads and triads do. Gender-based differences are observed as early as birth; newborn girls show a propensity to look at faces more than boys do.

Individuals' cultural backgrounds and the cultural context of social encounters also have strong effects on gaze production and perception. For instance, individuals from "contact" cultures such as Latin Americans look more at others than individuals from "non-contact" cultures such as the Japanese do. A consideration of cultural differences in gaze behavior is particularly important in the context of this dissertation as the first study was conducted in the United States with native American-English-speaking participants and the second and third studies were conducted in Japan with native Japanese-speaking participants.

Research has also found significant differences in gaze production across individuals even from the same gender and culture population due to differences such as personality type and tolerance for intimacy and arousal.

2.6. Chapter Summary

This chapter provided background on social gaze behavior from research on human communication, and related work on the design of social gaze behavior for embodied conversational agents and humanlike robots, focusing specifically on the social contexts described in the scenarios presented in the first chapter. The next three chapters will describe three empirical studies that are contextualized in and motivated by these scenarios. Each study involves a human-robot interaction situation where aspects of robot gaze play a central role. The next chapter will describe the first study, which focused on designing gaze cues for a humanlike robot to communicate attention and how these cues might lead to social improvements, particularly on learning and perceptions of the robot.

3. Study I: Designing Gaze Cues for Communicating Attention

This study investigated how robots might use gaze cues to communicate attention and whether effective communication of attention could deliver social and cognitive benefits. This understanding might have important implications for the educational use of humanlike robots, as described in the scenario provided in Section 1.1.1, particularly for understanding of whether robots can positively affect learning through behavior change. The results of this study also contribute to our understanding of the extent to which the findings from human-human communication (e.g., those found in Otteson & Otteson, 1980; Sherwood, 1987) carry over into contexts of human-robot interaction. Furthermore, the design of the gaze cues for this context of the study contributes to robot design by providing a computational model of gaze behavior and an understanding of how aspects of this model can be manipulated to obtain increased attention.

This investigation was contextualized in storytelling and, therefore, required gaining a better understanding of a human storyteller's use of gaze cues. Section 3.1 describes this step of theoretical and empirical grounding. The design specifications produced in this step informed the development of a computational model of gaze behavior, which was implemented on ASIMO for a storytelling application. Details of the

Figure 3.1. ASIMO, the humanlike robot used in this study.

computational model and the implementation are described in Section 3.2. Finally, an experimental situation in which the robot told stories to two participants was choreographed. The robot's gaze behavior was manipulated to favor (i.e., direct its gaze more towards) one of the two participants and the social and cognitive outcome of the robot's increased gaze was evaluated. Section 3.3 includes a description of the experimental evaluation. The findings and research implications of the study are discussed in Section 3.4.

3.1. Theoretically and Empirically Grounded Design

To provide a social context for the study, I designed a storytelling experience in which ASIMO (Figure 3.1), a humanoid robot developed by Honda (Sakagami et al., 2002), told the English translation of a Japanese fairytale, titled "The Tongue-Cut Sparrow" (Ozaki, 1970) to two listeners using a prerecorded voice. ASIMO's gaze behavior was designed based on a humanlike gaze model that used existing theory and empirical data collected from a human storyteller. Arm and body gestures were added to enrich the naturalness of ASIMO's behavior.

3.1.1. Theoretical Grounding

In this study, I extended the findings of Cassell et al. (1999b) who developed a predictive empirical model of human gaze behavior during turn-exchanges and within a turn based on the structure of the information conveyed by the speaker. Because the task involved storytelling, which follows the structure of an oratory instead of a conversation and does not involve turn-taking, the model was used to determine gaze shifts within the utterances of robot's speech (i.e., within a turn). Their model follows the English sentence structure suggested by Halliday (1967), who describes the two main structural components of an utterance using the terms *theme* and *rheme*. The theme refers to the part of an utterance that sets the tone of the utterance and connects the previous utterance to the next one. The rheme contains the new information that the utterance intends to communicate. For instance, in the sentence "The old women had made some starch." "The old women" is the theme while "had made some starch" is the rheme of the utterance. In the Cassell et al. model, speakers look away from their listeners at the beginning of a theme with 0.70 probability and look at their listeners at the beginning of a rheme with 0.73 probability. Figure 3.2 shows a graphical illustration of this model. Appendix A provides an algorithmic description.

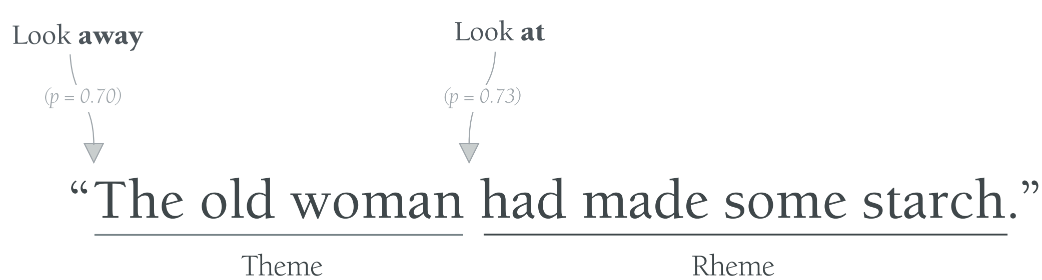


Figure 3.2. A representation of the gaze model suggested by Cassell et al. (1999b).

3.1.2. Empirical Grounding

Empirical data collected from a professional storyteller was used to determine locations and frequencies for the algorithm proposed by Cassell et al. (1999b). The professional storyteller was videotaped telling two stories to a two-person audience. Figure 3.3 illustrates the spatial configuration of the storyteller, the audience, and data collection equipment.

Thirty minutes of video data was used to analyze where in the environment and for how long each gaze shift executed by the storyteller was directed. The results showed that the storyteller gazed at four different kinds of locations: the two members of the audience, a fixed spot on the table in front of speaker, and a set of spots in the environment. Figure 3.4 shows a k-means clustering of these four locations.

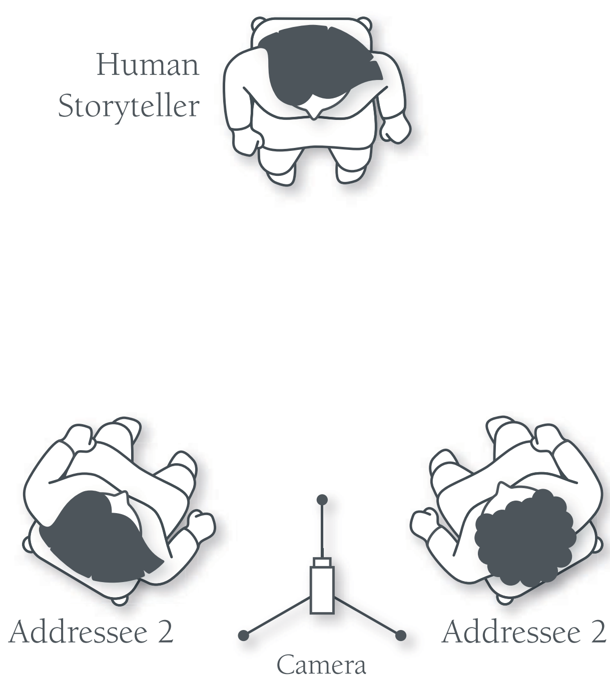


Figure 3.3. Th e spatial configuration of the data collection setup showing the storyteller, the audience, and the data collection equipment.

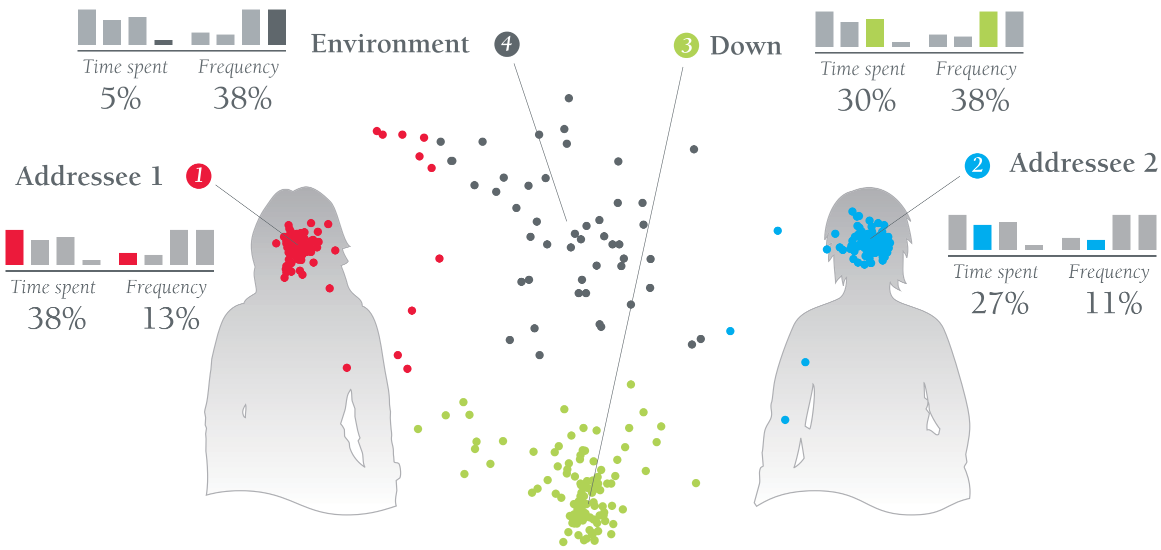


Figure 3.4. Th e gaze clusters identified through a k-means clustering and the total time spent looking and gaze shift frequencies for each cluster.

3.2. Implementation

ASIMO's gaze behaviors were designed by augmenting the Cassell et al. model with the gaze clusters and durations identified in the human storyteller gaze data. When gaze was currently directed at a listener, "looking at" was defined as keeping ASIMO's gaze on one of the two listeners. "Looking away" meant looking at the other listener or some other spot in the environment or down at a point of fixation. Moments during which the gaze was not directed at a listener, "looking at" meant looking at one of the listeners, while "looking away" meant looking at any four of the targets with predetermined probabilities. These probabilities were derived from an analysis of the frequencies of the human storyteller's gaze at each location. The duration of the gaze at each location was represented as a distribution, which was used to determine the length of the simulated gaze. These values are shown in Appendix A.

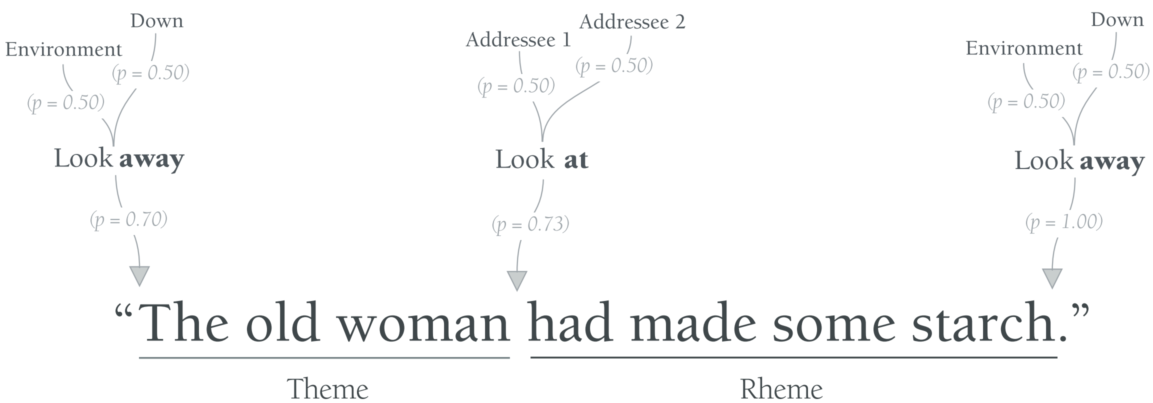


Figure 3.5. A representation of the gaze model developed by extending the model suggested by Cassell et al. (1999b) using empirical results.

This gaze model was used with a hand-coded script of the information structure of the fairytale to simulate humanlike gaze behavior, which marked the start of each theme and rheme and pauses between utterances. Figure 3.5 shows a representation of the gaze model developed by extending the algorithm proposed by Cassell et al. An algorithmic representation of the model is provided in Appendix B.

The gaze model was implemented on ASIMO by using a script of the story and synchronizing ASIMO's gaze behavior with a prerecorded voice. Ten simple arm gestures were automatically added for long utterances (greater than the mean length of 2,400 ms for gaze at a listener). Six special gestures such as bowing, crying, and acting angry were added by hand where they were semantically appropriate. The location of the participants was not sensed by ASIMO but was determined by placing two chairs at known locations and programming the robot to look in those two directions. The initiation of the robot's movement was controlled by the experimenter. The robot then introduced itself to the participants, told the story, and ended the interaction.

3.3. Experimental Evaluation

Drawing from existing theory on human-human interaction (Kleck & Nuessle, 1968; Cook & Smith, 1975; Fry & Smith, 1975; Otteson & Otteson, 1980; Sherwood, 1987), two hypotheses were formulated about responses to a manipulation in the amount of gaze behavior:

Hypothesis 1 – Participants who are looked at more by ASIMO will perform better in the recall task than participants who are looked at less.

Hypothesis 2 – Participants who are looked at more will evaluate the robot more positively than participants who are looked at less.

To test these hypotheses, a between-subjects experiment was conducted where participants listened to ASIMO while it told the Japanese fairytale in English. ASIMO's gaze behavior was manipulated so that its gaze was directed toward one of the

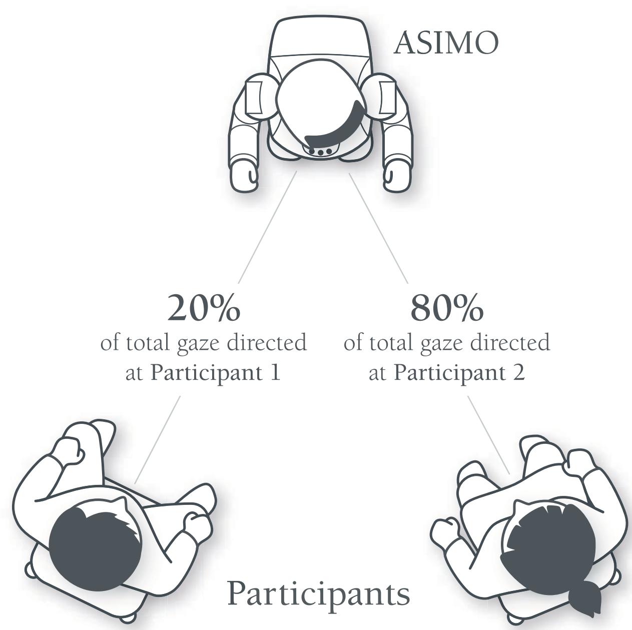


Figure 3.6. The spatial configuration of the experiment and description of the gaze manipulation.

participants with 20% frequency and the other participant with 80% frequency. This manipulation was done by changing the probabilities in the gaze model of looking at the two addressees from 0.50 and 0.50 (as in the human storyteller's gaze) to 0.80 and 0.20. Participants were placed at the same distance from ASIMO and space was left between them so that the robot's gaze toward each participant would be easily distinguishable. The experimental setup, shown in Figure 3.6, was similar to the configuration described by Otteson and Otteson (1980).

3.3.1. Experimental Procedure

Participants were first given a brief description of the experiment procedure. After the introduction, participants were asked to answer a pre-experiment questionnaire and were then provided with a more detailed description of the task. The participants were seated and ASIMO then introduced himself and performed the storytelling task (Figure 3.7). After listening to ASIMO's story, participants performed a distractor task,

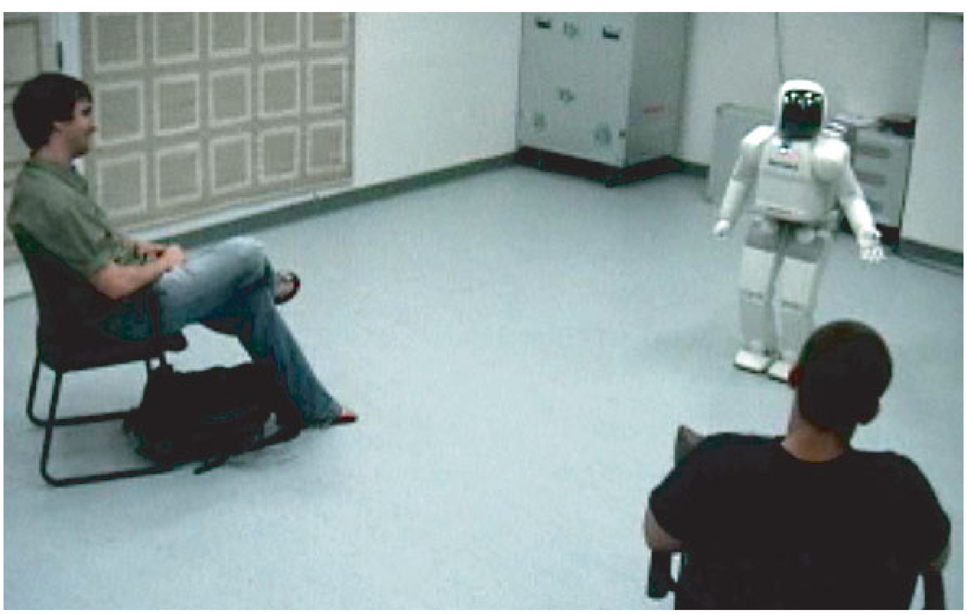


Figure 3.7. ASIMO telling its story to two participants in the experiment.

where they listened to another story on tape ("The Flying Trunk" by Andersen, 2001). Before listening to either story, they were told that they would be asked questions regarding one of the stories. All participants were asked questions regarding ASIMO's story. After completing the task, participants answered a post-experiment questionnaire regarding their affective state, perceptions of the robot, and demographic information. ASIMO's story, the story on tape, and the whole experiment took an average of 17.5 minutes, 7.5 minutes, and 35 minutes respectively. The experiment was run in a dedicated space with no outside distraction. A male and a female experimenter were present in the room during the experiment. All participants were paid $10 for their participation.

3.3.2. Measurement

All factors in the experiment were identical for each participant except for the two controlled factors: the frequency of the robot's gaze at each participant (a manipulated independent variable) and the participant's gender (a measured independent variable). The dependent variables measured were task performance, the participant's own affective state, their positive evaluation of the robot, their perceptions of the robot's physical, social, and intellectual characteristics, and their involvement in and enjoyment of the task. The post-experiment questionnaire included one question as a manipulation check, "How much did the robot look at you?" Seven-point Likert scales were used for all scales.

3.3.3. Participant Sample

Twenty undergraduate and graduate students (12 males, 8 females) from Carnegie Mellon University participated in the experiment. All participants were native English speakers and their ages ranged from 19 to 33 with an average of 22.6 (*SD*=3.72). Ten participants were assigned to the "looked at 80% of the time" condition. The other ten participants were assigned to the "looked at 20% of the time" condition. Participants were chosen so that they would represent a variety of academic majors including management sciences, social sciences, art, and engineering. Four male and three female participants had technical majors such as computer science, electrical engineering and information systems, while eight males and five females came from nontechnical fields including management and social sciences. On average, male participants had more video gaming experience (males; *M*=4.91, *SD*=2.21, females; *M*=3.13, *SD*=2.23) and more familiarity with robots (males; *M*=3.91, *SD*=1.76, females; *M*=2.75, *SD*=1.39) than female participants did.

3.3.4. Results

Three methods were used in the data analysis: repeated measures analysis of variance, analysis of variance (ANOVA), and multivariate correlations. The first method applied an Omnibus F-Test to see if the difference between pre-experiment and post-experiment measurements was significant across the two conditions and genders. The second technique used a linear regression (ordinary least squares estimation) on the variables that were significant across conditions to identify the direction of main effects and interactions. The last method looked at how these variables correlated with each other. Reliability tests and factor analyses were also conducted on the scales that were used for measurement.

Item reliabilities for all partner, task, and self-evaluation scales except the mutual liking scale (Cronbach's α=0.54) were high. However, because the scales for partner evaluation were created to evaluate humanlike interface agents, a factor analysis was

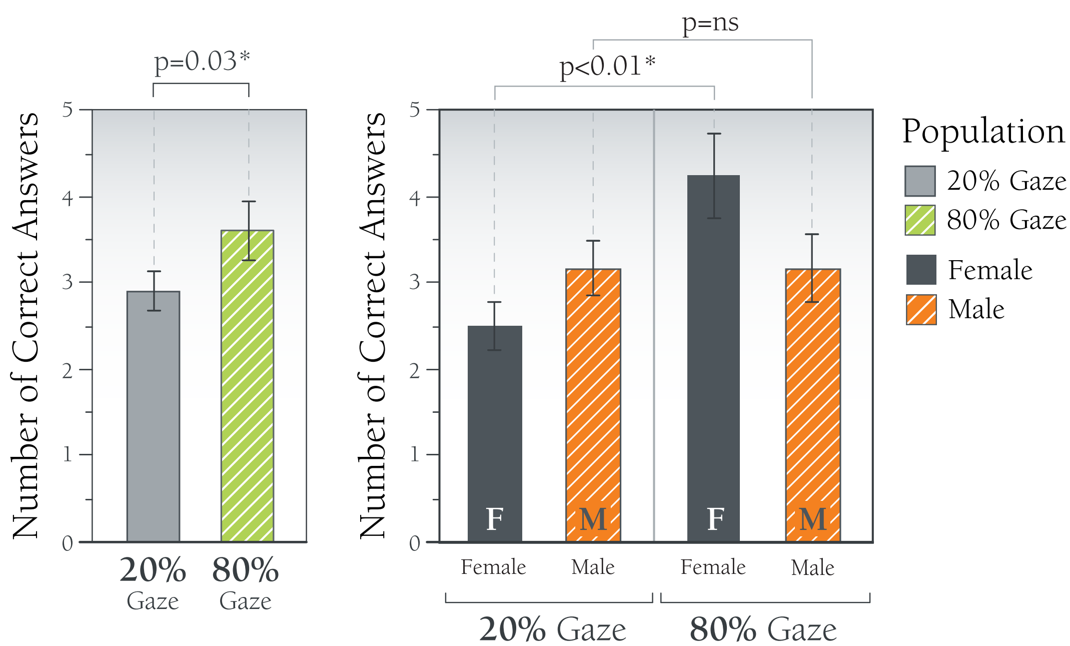


Figure 3.8. The number of correct answers in for participants in the 20% gaze and 80% conditions (left) and the breakdown of these results for female and males (right).

conducted including all the items that were used for evaluating the robot. This analysis created a highly reliable (Cronbach's α=0.91), 8-item scale for positive evaluation. An analysis of the manipulation check showed that the participants were aware that they were looked at more or less by the robot, $F(1,18)=4.29$, $p=0.05$.

Consistent with the first hypothesis, a regression on the performance measure showed that participants who were looked at more performed significantly better in the recall task (answering questions regarding ASIMO's story) than those who were looked at less, $F(1,16)=5.15$, $p=0.03$. When participant's genders were included in the statistical model, the effect was shown to be significant only in females ($F[1,16]=8.58$, $p<0.01$) there was no significant difference across conditions in males, $F(1,16)=0$, p=ns. These results are illustrated in Figure 3.8.

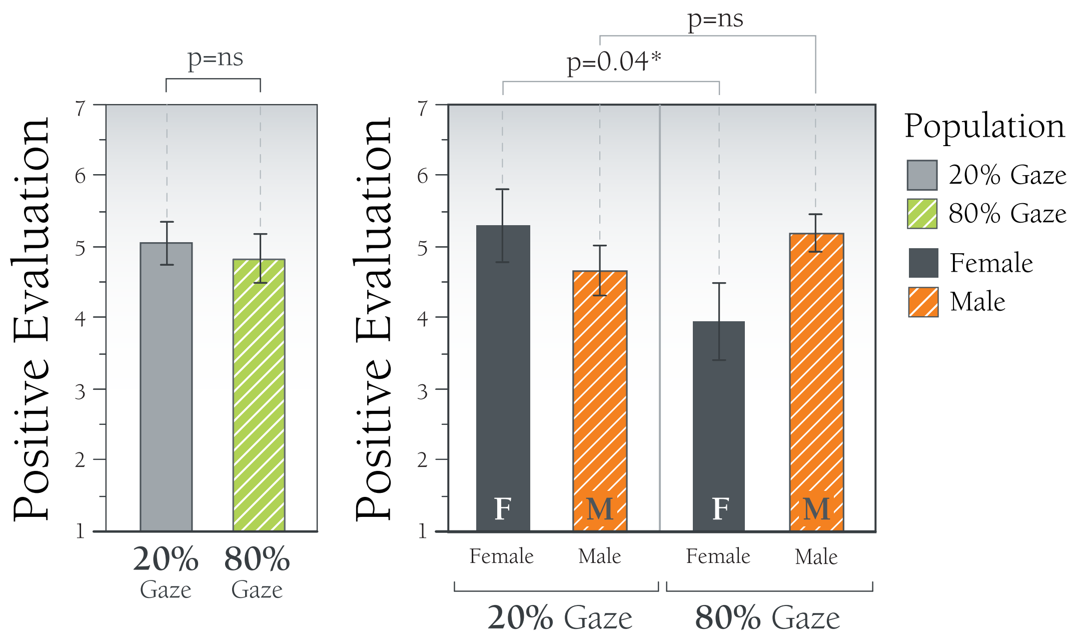


Figure 3.9. Positive evaluations of the robot in the 20% gaze and 80% conditions (left) and the breakdown of these results for female and males (right).

The analysis of the ratings of the positive evaluation scale of the robot showed no significant main effect but showed a significant interaction between experimental condition and participant gender, $F(1,16)=5.62$, $p=0.03$. Women rated ASIMO less positively when they were looked at more ($F[1,16]=4.80$, $p=0.04$), while men's evaluations of the robot did not significantly change based on how much gaze they received, $F(1,16)=1.14$, p=ns. These results are illustrated in Figure 3.9. Although this result reveals significant interactions with participant's gender, it is not consistent with the prediction in the second hypothesis. Analysis of scales of participantss affect, task enjoyment, and task involvement did not show any significant effects or interactions.

The last analysis looked at how the evaluation scales correlated with the participant's computer use, their familiarity with robots, and video gaming experience. A multivariate analysis using Pearson's correlation coefficient showed that ratings of the positive evaluation scale were highly correlated with video gaming experience (r=0.65, p<0.01), while not correlated with computer use or familiarity with robots. This correlation held for both genders although it was higher in males. Video gaming experience was also correlated with task enjoyment, r=0.53, p=0.02.

3.4. Discussion

The results supported the first hypothesis; the frequency of the robot's gaze affected performance on the recall task. This result suggests that, in the context of the first scenario described in Chapter 1, the teacher can instruct the robot to direct the attention of the distracted student to itself and its story, which will lead to better recall of the story. However, only female students might respond to the robot's increased gaze.

The second hypothesis, that participants who are looked at more will evaluate the robot more positively, was not supported. Furthermore, when gender was included as a variable in the analysis, women appeared to like the robot more when they were looked at less. This result might be explained by the differences in male and female perception of personal space based on the amount of mutual gaze established with a partner (Argyle & Dean, 1965; Hayduk, 1983). Bailenson et al. (2001) showed that these gender-based differences appeared in people's interactions with virtual agents. They found that female participants maintained more interpersonal distance between themselves and agents who built eye contact with them than with agents who did not. Male participants did not show similar changes in behavior. This finding might imply that because participants were not allowed to control their distance to the robot, females might have felt uncomfortable and evaluated the robot negatively when the robot gazed at them more. Lack of control over their distance with the robot did not affect males and they evaluated the robot more positively when it looked at them more.

An alternative explanation for the strong gender effects in information recall and evaluations of the robot is that men and women received different amounts of gaze due to differences in their average heights. As described earlier, ASIMO's gaze was directed at participants by seating them at known locations and programming the robot to look at these locations. However, this method did not control for the variability in participants' heights, particularly the differences in height between men and women (an average of 5.5 inches or 14 cm). Therefore, women might have received more direct gaze from the robot than men did in the experiment. However the manipulation check did not support this explanation; the analysis found no significant effect of gender on participants' ratings of how much gaze they received ($F[1,16]=1.57$, p=ns), or an interaction effect between gender and gaze manipulation over participants' ratings of how much gaze they received, $F(1,16)=0$, p=ns.

In correlation analyses, positive evaluations of ASIMO were found to be highly correlated with participants' video gaming experience and not with their computer use, which suggests that people might perceive ASIMO as more like a video game character or avatar than like a computer. This result suggests that people's interactions with robots and their interactions with embodied agents fall into the same ontological category of social responses to technology supporting the earlier argument that social responses to robots are evoked by humanlike cues instead of the automaticity of social behavior in the presence of minimal social cues.

3.4.1. Limitations

There are a number of limitations of this study, some of which are addressed in the second and third studies, which are discussed in the subsequent chapters. However, a number of limitations, particularly those imposed by the scope of the dissertation, the research approach, and the availability of resources, remain unaddressed. These limitations are discussed in Chapter 6.

The design of the gaze behavior did not account for some elements of the professional storyteller's gaze. For example, the speaker occasionally switched from looking at one listener to looking at the other listener during a theme or rheme, but the analysis did not identify a pattern in this behavior. The second study addresses this limitation by comparing gaze shifts across situations with one and two listeners.

While this study looked at how cues from a robot's gaze affected the communication of attention, other nonverbal elements such as arm gestures and postural changes were used to make the experience as fluid and natural as possible. However, some participants found ASIMO's arm gestures distracting, perhaps because of the servomotors that generate noise while moving the robot's arms and the slowness of the robot's arm motions relative to human motion. Furthermore, the robot used arm gestures based on its orientation of attention. When it looked at the participant on the right, it gestured with its right arm meaning that the robot directed more gestures beyond gaze at the person to whom it looked at more. This behavior might have heightened participants' feelings of being attended to by the robot. To remove this possible effect, arm gestures were eliminated in the next two studies.

Another limitation to the humanlikeness of ASIMO's gaze model was due to the physical design of the robot. When humans direct their gaze, their movements combine movement of the eyes, the head, and the upper torso, whereas ASIMO only used head movement to shift its gaze because its design does not include visible, controllable eyes and movement of the upper torso requires lifting and placing of the feet repeatedly, which were found to be time consuming and distracting in the pilot study. However, the results showed that this simple head movement was sufficient to create the experimental manipulation. Participants were asked to rate the amount of gaze they received from the robot. People who were looked at more thought the robot looked at them more (M=56, SD=19) and those who were looked at less thought ASIMO looked at them less (M=38, SD=20) with the difference being marginally significant, $F(1,18)=4.29$, $p=0.05$. While the results suggest that head movement is sufficient to evoke the feeling of being looked at, robots with visible, controllable eyes are used in the second and third study to achieve more natural humanlike gaze behavior.

Some participants found ASIMO's story, which was 17.5 minutes in length, to be too long. While it is important for participants to immerse into the experience with sufficiently long interactions, the second and third studies involve relatively shorter tasks.

A post-hoc analysis of the data on storyteller's gaze behavior showed that the length of gaze shifts followed a positively skewed distribution instead of a normal distribution. This caused ASIMO's gaze shifts to be longer on average than the human storyteller's gaze shifts. This error is corrected in the second and third studies by representing gaze lengths with two-parameter continuous distributions (specifically a Gamma distribution).

Finally, the results of this study are not fully conclusive because the experimental design did not include control (e.g., each participant receives 50% of the robot's attention) or baseline (e.g., participants listen to the story on tape or from a robot with stationary gaze) conditions.

3.5. Study Conclusions

This study investigated how a robot might use gaze cues to communicate attention and whether effective communication of attention by a robot can deliver social and cognitive benefits. The results showed that the robot could improve information recall simply by looking more at an individual. This finding has strong implications for educational use of robots and provides the first empirical evidence in human-robot interaction literature that robots can deliver cognitive benefits. The study also confirmed the causal relationship shown in human communication between increased gaze and improved information recall in the context of human-robot interaction, suggesting that results from human communication might carry over to human-robot interaction in the communication of attention.

The results also showed that the social and cognitive outcomes differed across male and female populations; only women's recall of information and subjective evaluations of the robot were affected by increased gaze.

The next study looks at how a robot might use gaze cues to shape the participant roles of its conversational partners and whether these roles evoke social and cognitive responses. It is described in detail in the following chapter.

4. Study II: Designing Gaze Cues for Signaling Conversational Roles

In human conversations, speakers use a number of cues that help to maintain fluent conversations, signaling the appropriate conversational roles to their partners. Gaze is an essential source for such cues and plays a significant role in conversational organization. An understanding of how robots might use these cues effectively to carry out fluent conversations with people not only impacts robot design, but also provides human-robot interaction research with a better understanding of how people respond to human conversational mechanisms when they are employed by robots. Furthermore, this exploration informs human communication research with computational models of conversational gaze behavior and a deeper understanding of the social and cognitive effects of conversational roles.

The human-robot interaction scenario described in Section 1.1.2 provides the context for this exploration. It starts with establishing a better understanding of how speakers use gaze cues to signal different participant roles, manage turn-exchanges, and signal the information structure by modeling the gaze behavior of a speaker in conversations with different participant role configurations. Detail is provided on this modeling in Section 4.1. The modeling is followed by building computational representations of a number of gaze mechanisms, combining these mechanisms into fluent conversational gaze behaviors, and implementing these behaviors on Robovie. Section 4.2 provides a description of the representation and implementation stages. Finally, a human-robot conversation scenario in which Robovie played the role of a travel agent and provided information to two participants is used to evaluate whether participants conformed to the conversational roles that the robot assigned to them through gaze and the social and cognitive outcomes that these roles evoked in participants.

4.1. Theoretically and Empirically Grounded Design

While the main focus of this study is to understand how robot gaze might signal participant roles in human-robot conversations, it necessarily takes into consideration those mechanisms essential for maintaining conversations. Research on conversational organization and linguistics suggest that gaze cues convey information about and are affected by two conversation-structural elements: "information structure"[1] (Cassell et al., 1999b) and "conversation structure"[2] (Kendon, 1967; Duncan, 1974; Sacks et al., 1974; Goodwin, 1981). Cassell et al. (1999b) found a tight temporal coupling between speaker's gaze shifts and points of transition in the information structure (the onset of the theme, the theme-rheme transition, and the end of the rheme) and conversation

[1] "Information structure" represents the variation in the sentence-level structure of how information is presented in the context of the emerging discourse (Roberts, 1996). This variation is roughly synonymous with parameters such as "theme/rheme" (Halliday, 1967) or "given/new" (Prince, 1981). The "theme" or "given" connects the utterance to the previous discourse and provides context for the new information to be presented. The "rheme" or "new" represents the new information that hearers could not have predicted from the context of the previous discourse.

[2] "Conversation structure" or "turn-taking" refers to the organization of speaking turns in a conversation that allows parties to hold the floor "one-party-at-a-time" (Goffman, 1955; Duncan, 1974; Sacks et al., 1974; Schegloff, 2000).

structure (turn-taking and turn-yielding). In the current design problem, gaze shifts were considered based on the following three conversation-structural elements:

> **Information Structure** – Natural, humanlike gaze shift patterns synchronized with the structure of the robot's discourse.
>
> **Conversation Structure** – Gaze cues for performing turn-exchanges (producing turn-yielding, turn-taking, and floor-holding gaze signals).
>
> **Participation Structure** – Gaze cues that signal participant roles.

To limit the scope of this investigation, the study focused on the following participant roles:

> **Addressees** – Participants who take speaking turns to contribute to the conversation and to whom the speaker addresses while speaking.
>
> **Bystanders** – Acknowledged non-participants who do not take speaking turns. The speaker does not address a bystander while speaking, but does acknowledge the bystander's presence during the conversation, particularly during greetings and leave-taking.
>
> **Overhearers** – Unacknowledged non-participants who do not take speaking turns. The speaker does not address an overhearer while speaking, and does not acknowledge the overhearer's presence at any point in the conversation. Here, it is important to note that the role of the overhearer was chosen to refer to the general category of unacknowledged non-participants for purposes of consistency. In the context of this study, this role is considered as interchangeable with eavesdropper or ignored.

4.1.1. Empirical Grounding Methodology

To better understand how people use the communicative mechanisms listed in the previous section, the study included formal observations guided by existing theory on conversational organization. These observations involved providing naïve participants with conversational scenarios involving different role structures, asking them to converse according to these roles, and observing how speakers used gaze cues to signal these roles. Participants conversed in the following role structures (illustrated in Figure 4.1):

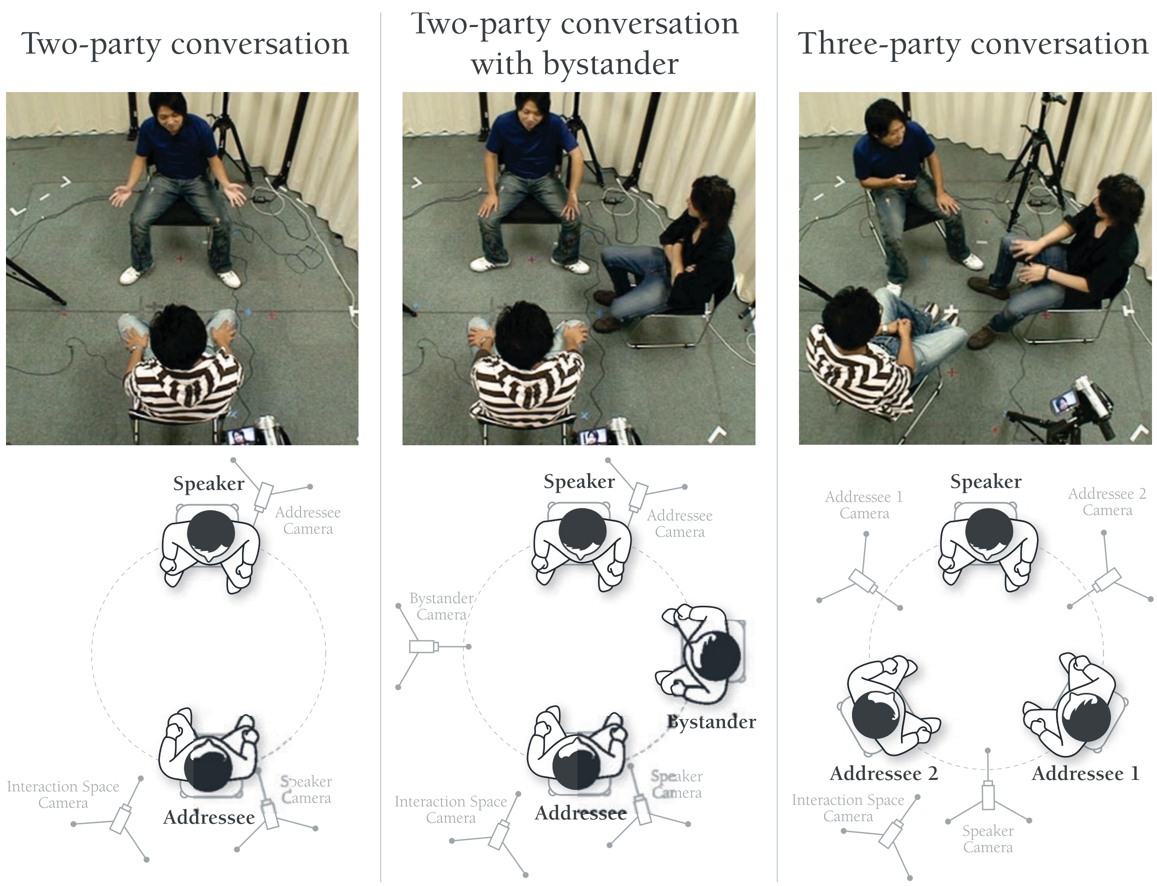


Figure 4.1. The data collection setup for the three conversational structures studied, a two-party conversation (left), a two-party conversation with a bystander (middle), and a three-party conversation.

Two-party conversation with a speaker, an addressee, and an overhearer:

> *Aiko (speaker) is a student at Osaka University and an active member of a student club. Takeo (addressee) is a new student at the same university. He is looking for a student club to join, therefore, attends the university club fair where he meets Aiko who is volunteering at her club's information booth. Aiko asks Takeo questions about his interests and provides him with information about club activities that might suit his interests.*

Two-party conversation with a speaker, an addressee, and a bystander:

> *Hiromi (speaker) is a resident of Shinsaibashi town in Osaka and attends Osaka University. He has lived in Osaka for a few years and is familiar with the town. Yoshi (addressee) is from Hokkaido and will be attending the same university. He is in town with his older brother Akira (bystander) to look for an apartment. A friend from high school connected him with Hiromi so that he can provide Yoshi with information about living in Osaka and local attractions that he might be interested in. As Hiromi tells Yoshi about the local attractions, Akira listens to their conversation.*

Three-party conversation with a speaker and two addressees:

> *Mika (speaker) is a student at Osaka University and works part-time at a local travel agency in Osaka. Toshi (addressee) and Jiro (addressee) are two friends attending Kinki University in Osaka. They plan to go on a trip together and are shopping for an affordable vacation package that they both find interesting. Mika inquires about their budget and shared interests and provides information about his company's travel packages.*

4.1.1.1. Participant Sample

Four all-male triads (12 participants) performed in the three scenarios described above. All of the participants were university students from the Osaka area of Japan whose ages ranged from 18 to 24. The experimental procedure was as follows: Before their participation, all participants were asked to review and sign consent forms. Next, they were asked to provide demographic information and fill in a questionnaire that measured their levels of introversion-extroversion (Goldberg et al., 2006). All triads performed in each scenario for fifteen minutes in the order that the scenarios are listed above: a two-party conversation, a two-party conversation with bystander, and a three-party conversation. At the beginning of each scenario, participants were provided with a description of the scenario and their roles. They had five minutes to ask questions and adapt to their roles. Between each pair of scenarios, participants were asked to solve ten-minute-long crossword puzzles to distract them from their roles in the previous scenario. At the end of their participation, each participant received ¥1,500 (approximately $14) in compensation.

4.1.1.2. Measurement and Selection of Data for Detailed Analysis

High-definition cameras placed across from participants' seats captured participants' gaze behaviors and stereo microphones attached to their collars captured their speech. The cameras provided video sequences of participants' faces (from hair to chin). An additional camera placed on the ceiling captured the interaction space. For both practical and ethical reasons, cameras remained visible to the participants. In total, the four cameras captured approximately 45 minutes of video for each participant (180 minutes of data in total for each triad.

An important limitation of modeling the behavior of naïve participants' role-playing in predetermined scenarios is the low fluency of interaction led by factors such as certain aspects of personality (e.g., introversion), experimenter effects (e.g., performance anxiety), and unfamiliarity with the topic of the scenario and with conversational partners. In order to base the behavioral model on data that is minimally affected by these factors, data from the triad that exhibited the most fluent interaction[3] was used in the detailed analysis. To limit the focus to speakers' use of gaze cues for signaling participant roles, the analysis included only the speaker's gaze behavior. However, because certain conversational mechanisms such as turn-taking necessarily involve reciprocity, it also considered addressee gaze and speech, but only at turn-exchanges. For purposes of simplicity, interruptions and backchannel responses—short utterances, such as "uh-huh," "yeah," and "okay," produced by one conversational participant while the other is talking (Ward & Tsukahara, 2000)—were omitted in the analysis.

4.1.2. Design of Conversational Mechanisms

The designed gaze behavior used an understanding of the basic spatial and temporal parameters of gaze cues and aspects of three conversational mechanisms that signal information, conversation, and participation structures. This section describes the results of the analysis for each mechanism.

[3] As evidenced by a qualitative evaluation of participants' involvement in the conversation and quantitative assessments of the least total time spent speaking with substantial pauses.

4.1.2.1. Spatial and Temporal Parameters of Gaze Cues

To identify the spatial and temporal characteristics, the analysis of the speaker's gaze behavior attempted to answer the following questions. Where do speakers look in different participation structures? How much time do they spend looking at these targets? The answers that the analysis provided directly informed the design of the robot's gaze behavior. The results of the analysis and elements of the design are described below.

Where Do Speakers Look?

A frame-by-frame analysis of the speaker's gaze behavior provided an understanding of the gaze targets. The analysis included coding the target and time of execution of each gaze shift by taking the speaker's perspective to qualitatively estimate the gaze targets and marking them on an image representation of the speaker's field of view. Next, gaze targets were clustered both qualitatively and quantitatively. The qualitative analysis identified three clusters in the two-party conversation and five clusters in the two-party conversation with bystander and the three-party conversation. The quantitative analysis used a Gaussian Mixture Model (GMM) estimation algorithm (Bouman, 1997) to determine the number of clusters and identify the cluster to which each gaze target belonged. This analysis confirmed the numbers of clusters in the qualitative assessment of the first and third scenarios. However, the clustering algorithm identified eight clusters in the second scenario, four of which included gaze shifts away from conversational partners and towards the environment. These four clusters were combined to make up a single cluster.

The three clusters identified in the two-party scenario corresponded to (1) the addressee's face, (2) the area of the addressee's body, and (3) spots in the environment, mainly aligned with or above the level of addressee's eyes. The five clusters identified in the two-party-with-bystander scenario corresponded to (1) the addressee's face, (2) the area of the addressee's body, (3) the bystander's face, (4) the area of the bystander's body, and (5) spots in the environment, mainly aligned with or above the level of the addressee's eyes. The five clusters identified in the three-party scenario corresponded to (1) the first addressee's face, (2) the area of the first addressee's body, (3) the second addressee's face, (4) the area of the second addressee's body, and (5) spots in the environment, mainly the area between the two addressees. Figure 4.2 illustrates gaze locations and identified clusters.

An inter-coder reliability analysis assessed the objectivity of the primary coder's analysis of the speaker's gaze direction. This analysis asked both the primary and the secondary coder to categorize a randomly selected sample from the video data (90 clips with single gaze shifts, 30 for each conversational scenario) into the clusters identified by the qualitative and quantitative assessment. A Cohen's Kappa (κ) of 0.78 was calculated to measure the agreement between the two raters, indicating substantial agreement. Disagreements between the two raters were resolved through discussion.

How Much Time Do Speakers Spend Looking at Each Target?

In all scenarios, the speaker looked at the addressee(s) for the majority of the time—74%, 76%, and 71% in the two-party, two-party-with-bystander, and three-party scenarios respectively. However, in the first two scenarios, the speaker looked toward the bodies of the addressees more than their faces (26% and 25% at the faces and 48% and 51% at the bodies). I speculate that this behavior is a form of intimacy regulation —to reduce the arousal increased by establishing eye contact with a conversational

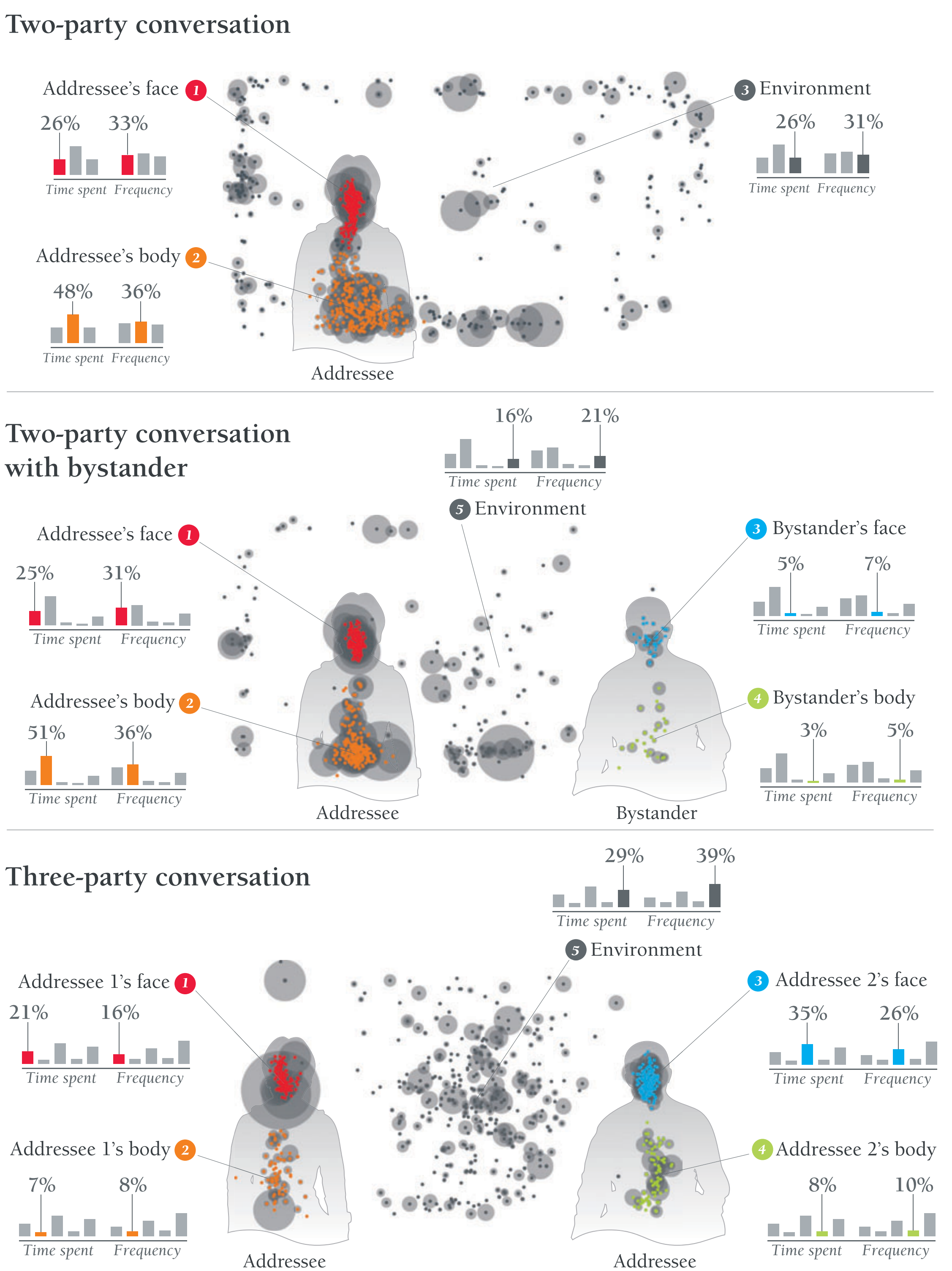


Figure 4.2. The gaze clusters and how much and how frequently these clusters are looked at for the three conversational structures studied.

partner (Argyle & Dean, 1965). I also observed that the gaze shifts towards the body often followed shifts towards the face, providing further support for the intimacy-regulation explanation. This behavior was less prominent in the third scenario (56% at the faces of the two addressees and 15% at their bodies), perhaps because shifting gaze to look towards another interlocutor serves as an alternative way of regulating intimacy (similar results were obtained by Vertegaal et al., 2001). Gaze breaking (i.e., avoiding eye contact) was also a common behavior that was observed in the video data. In all scenarios, the speaker spent a significant amount of time looking away from addressees (26%, 16%, and 29% of the time in the three conversational situations respectively). These findings are summarized in a graphical form in Figure 4.2.

A qualitative assessment of the duration of gaze shifts at each target showed that these durations could best be represented with two-parameter continuous distributions. Therefore, estimates for these parameters were calculated by fitting Gamma distributions (defined by parameters θ for shape and k for scale) to the data from each gaze cluster. Appendix C provides distribution parameters for each gaze cluster.

4.1.2.2. Gaze Cues that Signal Information Structure

As discussed in the first study, research in computational linguistics suggests that information structure—the relationship between the context of utterances or clauses and the emerging discourse context—accounts for a large portion of speaker gaze shifts within the course of a turn (Cassell et al., 1999b). In the first study, a model developed by Cassell et al. (1999b) was developed to generate gaze shifts that signaled the information structure of the robot's utterances. However, because the linguistic context of the second study is the Japanese language and the relationship between gaze and information structure in Japanese has not been extensively studied, I instead sought to model this relationship from empirical data.

I began by identifying the most appropriate unit of analysis for the speaker's speech. To model the relationship between gaze and information structure in English, Cassell et al. used "utterances" (1999b), also called "idea units" (Chafe, 1993), as the unit of spoken discourse for their analysis. However, Maynard (1989) argued that the highly fragmented structure of the Japanese spoken language could better be modeled using a smaller unit of analysis, called Pause-bounded Phrasal Units (PPU). Alternatively, spoken discourse that follows the conversational style of a "casual narrative" (Tannen, 1984) in which the speaker holds the floor for longer periods can be modeled for thematic discourse segments (Hinds, 1976; Maynard, 1989) that perform as pragmatically functional speech acts and cause shifts in the participants' points of view (Maynard, 1989). Because the conversational style of the data collected for this study had the characteristics of a casual narrative, I chose to use thematic segments to model the speech data.

Thematic segments, also called "paragraphs" of a discourse by Hinds (1976) or "thematic fields" by Maynard (1989), represent distinct discourse topics, consist of sentences that are more closely related to each other than to other sentences in the discourse, and are bounded by topic shifts marked by linguistic and interactional expressions. The following features characterize these expressions:

> **Substantial lapse, often filled with back-channel-like utterances –** The speaker pauses as a result of hesitation that might signal a move from one focus to another or from one thought to another (Chafe, 1979).

Formulation and evaluative comments – The speaker uses a part of the conversation to summarize, characterize, translate, or explain what is being said (Garfinkel & Sacks, 1970).

Minimal responses – Responses such as acknowledgements, mirror responses, and laughter that do not contribute to the advancement of the topic and are often followed by pauses (McLaughlin & Cody, 1982). In Japanese, minimal responses that mark the transition of a thematic field are conclusive remarks that are pronounced with a finalizing tone such as "naruhodo ne," "I see" (Maynard, 1989).

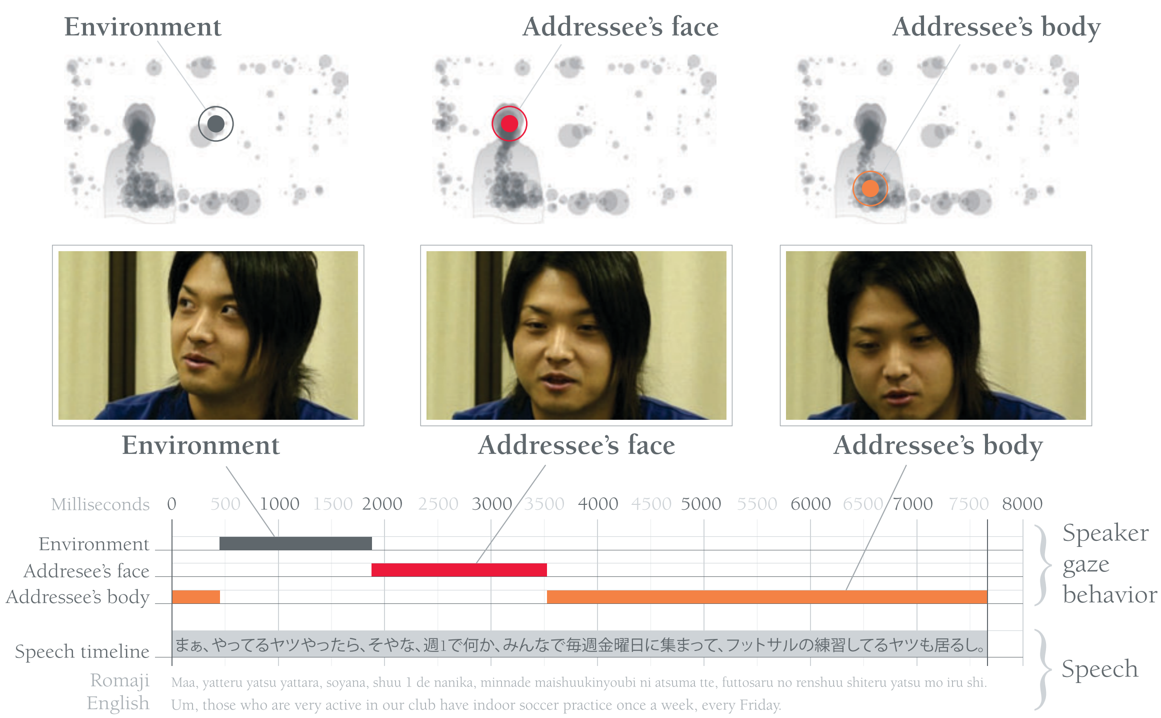


Figure 4.3. One of the gaze patterns that signals information structure identified in the two-party conversation.

Sentence adverbs and conjunctions – Transitional adverbs and conjunctions that fill the gap between two themes to minimize disruptions in conversational flow such as "tokorode," "by the way" (Maynard, 1989).

Following Maynard's (1989) description, a native Japanese speaker coder unitized the speech data into thematic fields, producing 181, 146, and 155 units in the three conversational scenarios respectively. Using a data visualization tool that I developed for this analysis, I mapped each thematic field onto the speech timeline along with gaze shifts that took place within the thematic field and 4000-millisecond periods before the beginning and after the end of the thematic field. This mapping allowed me to qualitatively identify patterns in gaze shifts that occurred at the onset of each thematic field and quantify the frequency of occurrence for each pattern. The analysis identified two main recurring patterns of gaze shifts in the two-party conversation and the two-party-with-bystander conversation and another set of two patterns in the three-party conversation. Figure 4.3 provides a graphical illustration of one of the patterns identified in the two-party conversation. The other patterns and the frequencies of each pattern are included in Appendix D.

4.1.2.3. Gaze Cues that Signal Conversation Structure (Turn-Taking)

Research in conversational organization and nonverbal behavior has shown that gaze behavior is also instrumental in managing turns in conversations and follows a common pattern (Kendon, 1967; Duncan, 1974; Sacks et al., 1974; Goodwin, 1981). To identify how gaze cues facilitated turn-exchanges, I identified the "turn-relevant places" in the speech data (where a speaker passed the floor to another speaker) based on the set of rules suggested by Sacks et al. (1974) and analyzed speaker's gaze direction during these exchanges. The analysis focused on turns that the speaker initiated with an explicit "turn-yielding" signal (as described by Duncan, 1974), omitting interruptions and overlapping speech. The analysis showed that three types of turn-management signals (also proposed by Kendon, 1967 and Duncan, 1974 for conversations in English) facilitated all of the turns that did not involve interruptions or simultaneous speech:

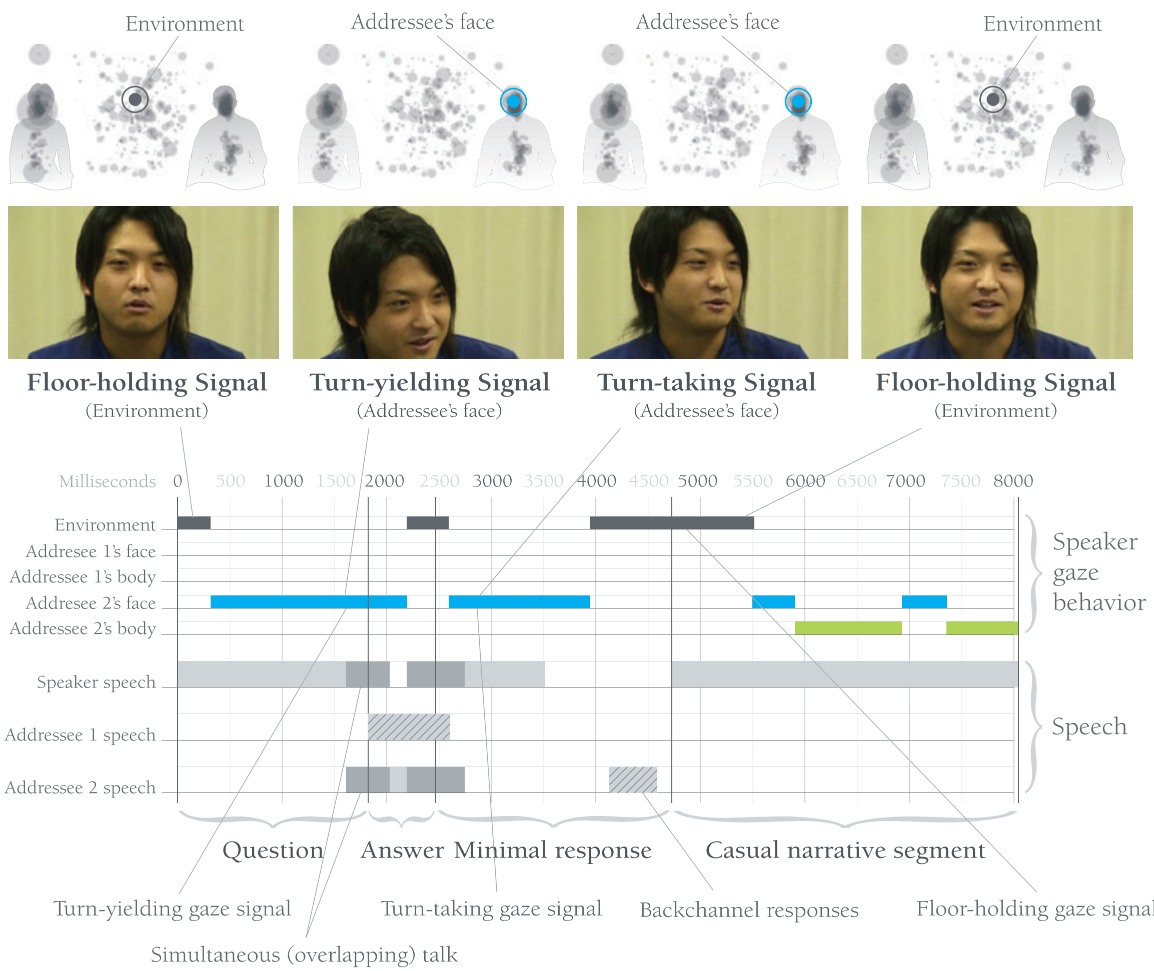


Figure 4.4. The sequence of four signals that the speaker used to manage turn-exchanges: floor-holding, turn-yielding, turn-taking, and floor-holding signals.

Turn-yielding – The speaker looks toward the addressee at the end of a turn accompanied by an evaluative remark or question, signaling to the addressee that the speaker is ready to pass the floor to the addressee.

Turn-taking – The addressee looks at the speaker at the end of the speaker's turn, signaling to the speaker that the addressee is open to taking the floor.

Floor-holding – As the next speaker takes the turn, the new speaker looks away from the new addressee, signaling holding the floor until the turn is complete.

"Question-answer pairs" are commonly observed cases of turn-exchanges. In these sequences, the speaker (1) produces a turn-yielding signal at the end of a question, (2) looks continuously at Addressee 1 during the response, (3) when Addressee 1's response is complete, the speaker produces a minimal response (McLaughlin & Cody, 1982) such as an acknowledgement, mirror response, or laughter, during which the speaker looks at Addressee 1, and finally (4) produces a floor-holding signal by looking away from Addressee 1 when starting casual speech. The analysis identified a total of 8, 9, and 20 question-answer pairs in the three conversational scenarios studied respectively. Figure 4.4 illustrates the speaker's turn-yielding, turn-taking, and floor-holding gaze signals during a question-answer pair that were observed in the data.

4.1.2.4. Gaze Cues that Signal Footing Structure (Participant Roles)

The analysis of role-signaling gaze cues showed that the speaker produced four sets of gaze cues to signal to the interlocutors their participant roles. These cues are described

Gaze target	During greeting	After greeting	During turn-exchanges	During speech
Addressee	**Gaze** (acknowledging gaze)	**Gaze**	**Gaze** (turn-yielding gaze)	**Gaze**
Bystander	**Gaze** (acknowledging gaze)	No gaze	No gaze	**Gaze** (short, acknowledging glances)
Overhearer	No gaze	No gaze	No gaze	No gaze

Table 4.1. A summary of the speaker's role-signaling gaze cues identified in the analysis.

below and are categorized based on where in the conversation they occurred. A summary of these cues is also provided in Table 4.1.

Greetings and summonses – An important point where speakers signal the roles of their conversational partners (and others signal their availability for these roles) is the opening of a conversation such as greetings where participants welcome and acknowledge each other, or summonses where one participant attracts the attention of another to start a conversation. Goffman (1955) describes greetings as serving "to clarify and fix the roles that participants will take during the occasion of the talk and to commit participants to these roles." Bales (1951, 1970) suggests that speakers rely primarily on gaze cues to signal these roles. Schegloff (1968) depicts an observation where the lack of gaze cues during a summons leads to ambiguity in who is being addressed in a crowd of bus-riders. In the observation, the speaker greeted and looked toward individuals in the roles of both addressee and bystander.

Transition from greetings to the body of the conversation – In the second conversational scenario, at the point of transition from greetings to the body of the conversation, the speaker diverted gaze *toward* the addressee and *away* from the bystander, providing a significant cue for their participant roles. In contrast, the speaker produced subsequent glances at both interlocutors in the three-party conversation, signaling to them their roles as addressees.

The body of the conversation – The speaker spent the majority of the speaking time looking at addressees. In the first conversational scenario, the speaker looked toward the addressee 74% of the time and the environment 26% of the time. In the second scenario, the speaker allocated some of the gaze toward the bystander (8%), mostly in short acknowledging glances averaging nearly half the average length of the gazes at the addressee (in seconds, M=0.77, SD=0.58 vs. M=1.40, SD=1.30). The speaker looked toward the addressee, bystander, and the environment 76%, 8%, and 16% of the time respectively. Finally, in the last scenario, the speaker looked toward the addressees and the environment at 71% and 29% of the time respectively.

Turn-exchanges – Another important point in conversations where participant roles are renegotiated is turn-exchanges. For instance, Weisbrod (1967 as described in Kendon, 1967) observed in seven-party conversations that the person at whom the speaker looked at the end of a turn was more likely to take the next speaking turn. In my observation, addressees received all turn-yielding gaze signals and bystanders received none, suggesting that the turn-yielding gaze cues are also important signals for establishing the footing structure of a conversation. In the three-party conversation, after the greeting, the speaker divided attention between the two addressees, switching gaze direction from one addressee to the other and waiting for one of the addressees to take the floor. Once the floor was taken, the conversation roughly followed the pattern of a sequence of two-party conversations. The speaker addressed and looked mostly toward one of the addressees and switched focus when

Figure 4.5. Robovie R-2, the humanlike robot used in this study.

the other addressee interrupted with an attempt to take the floor, when a question was directed at both addressees and was answered by the other addressee, or at points of significant shift in the topic of the conversation.

4.2. Implementation of the Design Elements

The findings from these analyses are represented as a hierarchical probabilistic state machine and implemented on a robotic platform that embodied the physical characteristics required by the behavioral model. The robotic platform used in this study was Robovie R-2 (Figure 4.5), a humanoid robot developed by ATR (Ishiguro et al., 2001).

Spatial and Temporal Parameters – The gaze model used the results from the analysis of the spatial and temporal parameters of the gaze cues to determine the robot's gaze direction in the three conversational situations described above, a two-party conversation, a two-party conversation with bystander, and a three-party conversation. Gaze target clusters were represented as two-dimensional normal distributions defined by their centers and spreads in gaze rotation angles (in degrees). The exact target of each gaze shift was randomly generated using the parameters of these normal

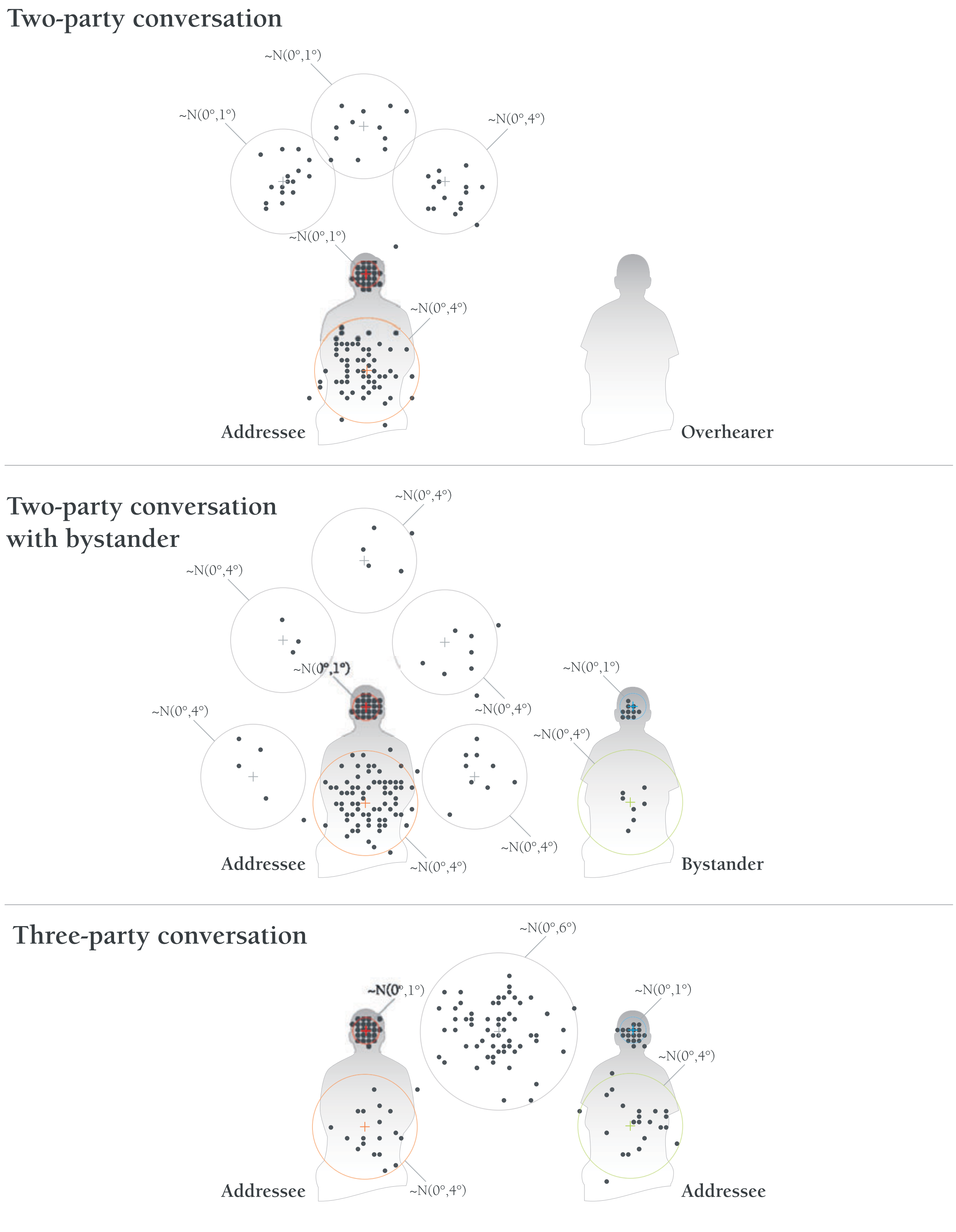


Figure 4.6. Gaze targets generated by the robot following the model created for the three conversational structures studied.

distributions. The model used the gaze duration distribution parameters calculated for each cluster to determine how long the robot should spend looking at each target. Figure 4.6 shows the gaze targets generated by the robot for different conversational situations.

The gaze shifts of the robot were divided into eye and head movements with a 1:1 vertical ratio and a 4:1 horizontal ratio. These ratios were determined based on the robot's pan and tilt ranges, motor speeds, and smoothness of motion to optimize for speed of gaze shifts and naturalness of the behavior. Also, each eye was given the appropriate horizontal angle for convergence (e.g., 1.5 degrees when looking toward a conversational partner at a two-meter distance).

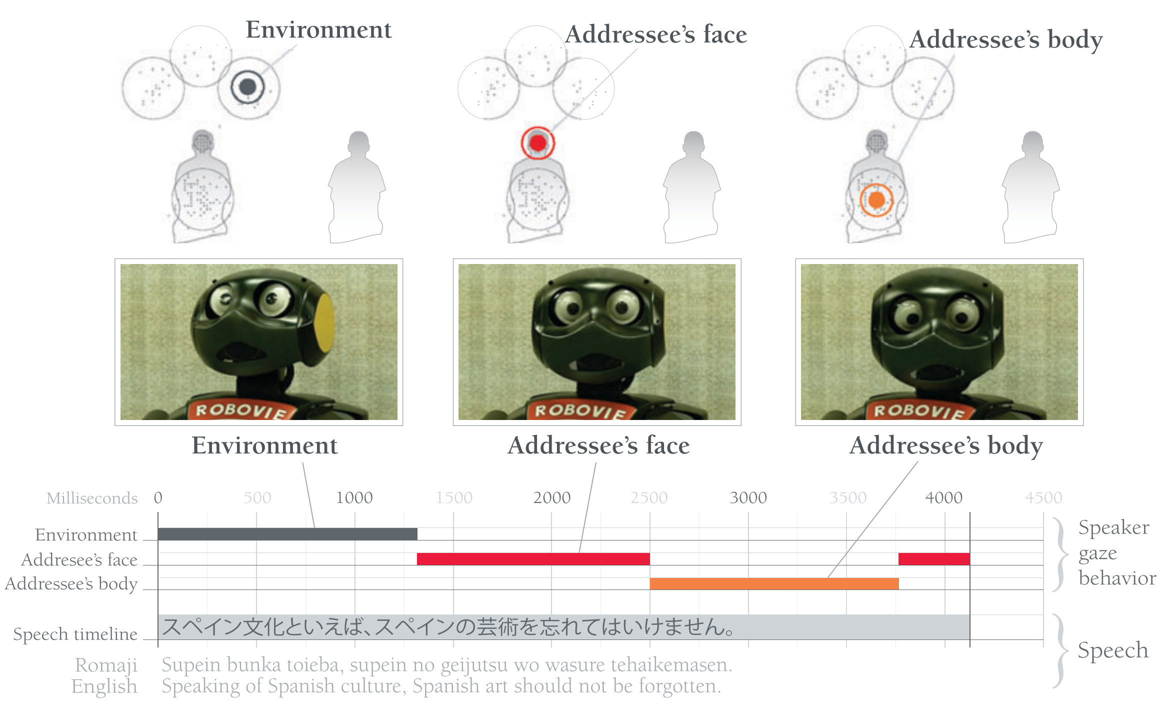


Figure 4.7. The robot producing one of the gaze patterns identified in the two-party conversation.

Signaling Information Structure – The gaze model also followed the patterns that were induced by the information structure of the speaker's speech. The robot's speech was marked for thematic fields. For each new thematic field, the robot produced the appropriate gaze pattern based on the probability of occurrence for the gaze pattern and calculated the length of the gaze shifts in the pattern based on the length distributions for each gaze. Figure 4.7 illustrates one of the frequent patterns that were identified in the two-party conversation as it was produced by the robot.

Signaling Conversation Structure – The robot's speech was also marked for turn-relevant places. During turn-exchanges, the robot looked at its addressee at the end of a question, producing a *turn-yielding* gaze signal. Towards the end of its partner's

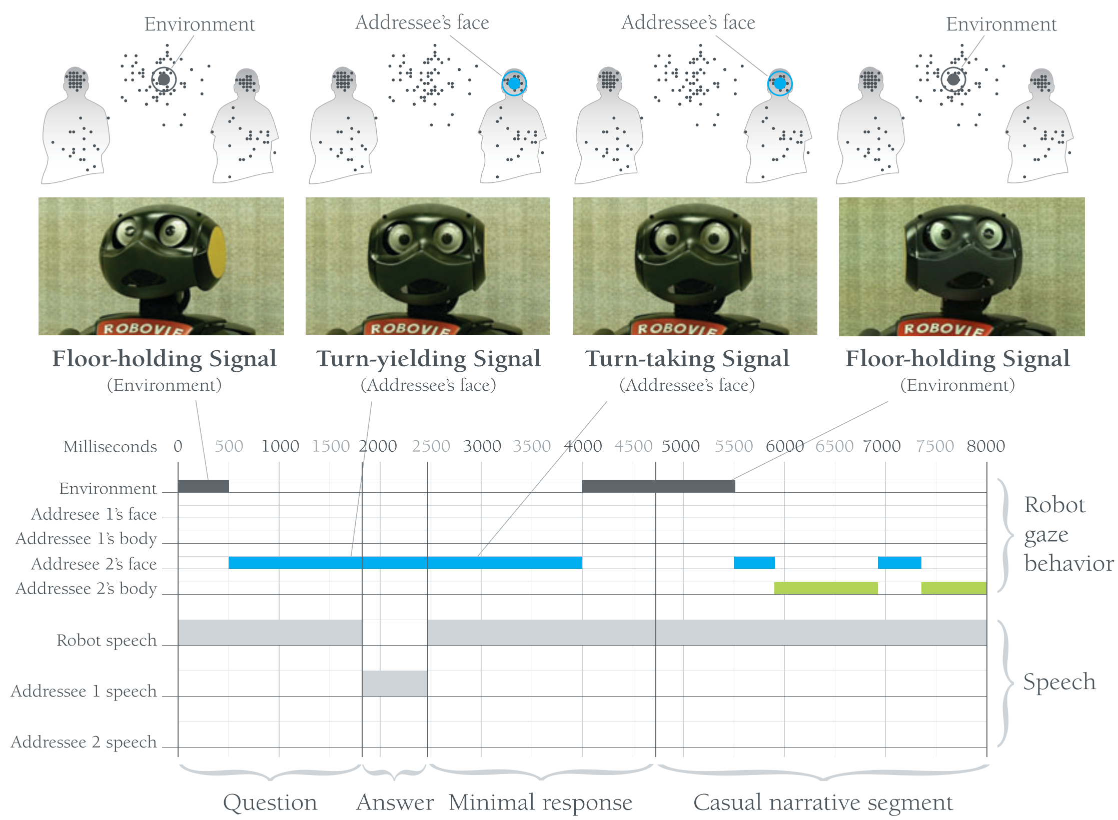


Figure 4.8. The robot producing the four subsequent signals that help manage turn-exchanges.

response, it looked at its partner, producing a *turn-taking* signal. Finally, when it took the turn, it looked away from its partner, producing a *floor-holding* signal. During "minimal responses," the robot looked at its addressee as the human speaker did in question-answer pairs. Figure 4.8 shows Robovie producing these signals in a question-answer pair.

Signaling Participation Structure – The robot's gaze behavior adapted to the three conversational scenarios constructed for modeling speaker gaze behavior. In the two-party conversation, the robot acknowledged its addressee during greeting and leave-taking, spent most of its time looking at the addressee's face or body (following the patterns that the analysis identified in the two-party conversation) and produced turn-yielding signals for the addressee. In the two-party conversation with the bystander, in addition to the behaviors it produced for in the previous conversational scenario, it greeted the bystander during greetings and leave-taking and reaffirmed the bystander's role with short glances directed at him at random intervals during the body of the

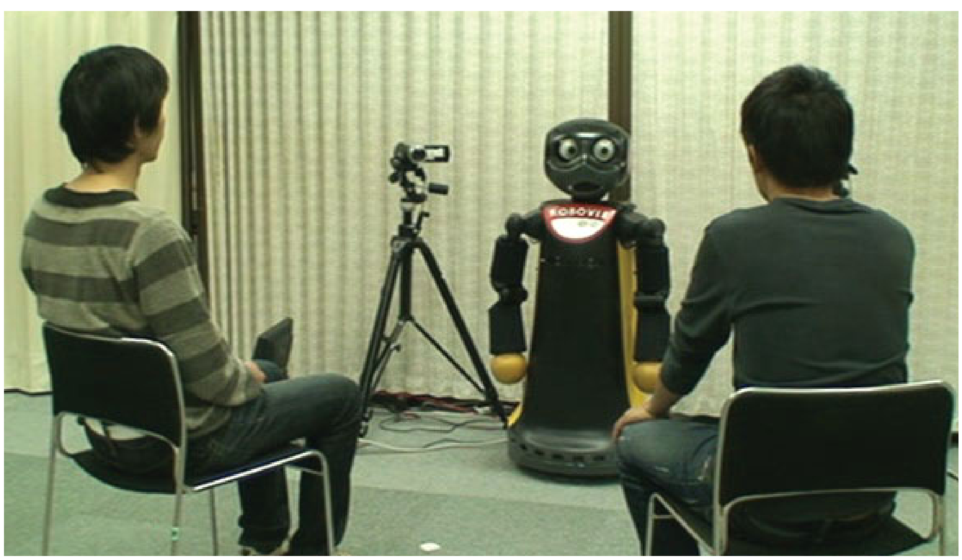


Figure 4.9. Participants conversing with Robovie in a conversational scenario.

conversation. Finally, in the three-party conversation, the robot greeted both addressees during greeting and leave-taking, looked at both of them during the body of the conversation following the patterns that the analysis identified in the three-party conversation, and produced turn-yielding signals for both partners.

Appendix E provides a summary of all the gaze cues that the robot produced to signal participation structure (i.e., footing), conversation structure (i.e., turn-exchanges), and information structure (i.e., thematic fields).

4.3. Experimental Evaluation

A controlled laboratory study in which naïve participants were asked to converse with a robot (as seen in Figure 4.9) in different participation structures evaluated how the designed gaze mechanisms might structure participant roles in human-robot conversations. The evaluation attempted to find answers to the following questions: Does a robot's use of designed gaze cues create different types of participation structures in a conversation? Do people conform to their participant roles? Would different participant roles lead to significant social outcomes such as stronger feelings of groupness or more liking of the robot? This section describes the experimental design, hypotheses, experimental procedure, measures, participant profile, and results of the experiment.

4.3.1. Experimental Design

To contextualize the design of gaze mechanisms for the robot, I choreographed a conversational scenario in which Robovie played the role of a travel agent. The robot provided participants with options of travel packages (value and premium) and destinations (Spain and Turkey) and adapted the information it provided to their choices. It also assessed participants' knowledge of the travel destinations that they chose by asking them factual questions, such as "Are you familiar with Picasso?" or "Did you know that Spain is this year's World Champion in basketball?" Wizard-of-Oz techniques were used to process participants' responses to the robot's questions and remarks. Below is a typical question-answer pair from the experiment:

> *Robovie:* *[Looking toward one of the participants] Did you know that the world's first coffee shop opened in Istanbul in the 15th century?*
>
> *Participant:* *Oh, I didn't know that.*

The robot followed the common interaction rituals of a conversation (as described in Goffman, 1971; Forgas, 1979). During greeting, the robot introduced itself to its conversational partners, asked them for their names, and told them that it was happy to meet them. During leave-taking, the robot told its partners that it had to talk to

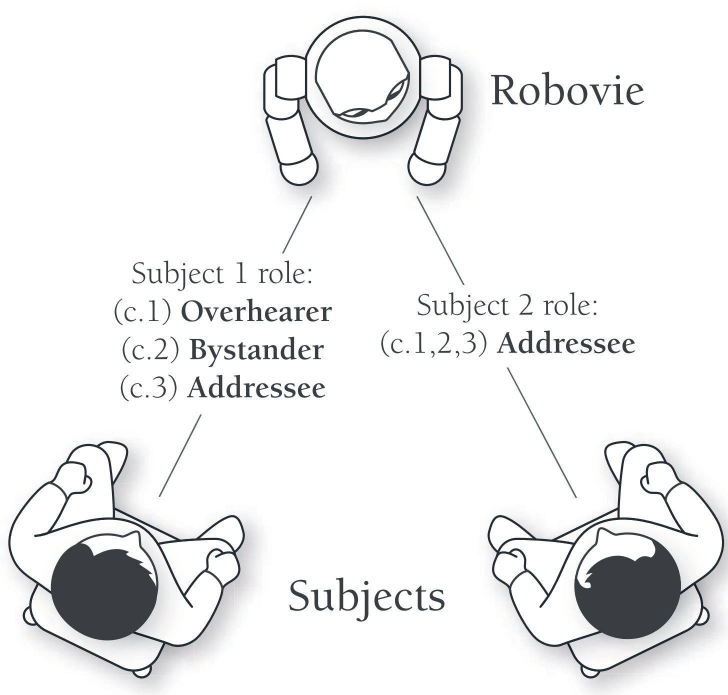


Figure 4.10. The spatial configuration of the experiment and description of the gaze manipulation.

another customer, but that it was nice meeting them, and thanked them for their interest.

Robovie's speech was identical across conditions except for changes due to the adaptive dialog. A prerecorded gender-ambiguous voice was used for Robovie's speech. Speech recognition was not used during the experiment. Instead, an experimenter initiated the robot's turns in the conversation, selecting from among a preset sequence of utterances from a library. Following a between-participants design, the robot's gaze behavior was manipulated in three conditions:

Condition 1 – The robot produced gaze cues for an addressee and an overhearer (ignoring the individual in the latter role).

Condition 2 – Gaze cues were produced for an addressee and a bystander.

Condition 3 – The robot produced gaze cues for two addressees.

Figure 4.10 illustrates the spatial configuration of the robot and subjects.

4.3.2. Hypotheses

Four hypotheses were developed from existing theory on conversational participation in human-human interaction, person perception, and group formation. To distinguish conversation participants (those who participate in a conversation by taking speaking turns) from experiment participants (those who were recruited to participate in the experiment), the latter will hereafter be referred to as "subjects."

Hypothesis 1 – Subjects will correctly interpret the footing signals that the robot communicates to them and conform to these roles in their participation to the conversation. Therefore, those who are granted speaking turns (addressees) by the robot will take more speaking turns and speak longer than those who are not granted speaking turns (bystanders and overhearers).

Hypothesis 2 – Subjects who contribute to the conversation by taking speaking turns (addressees) will have better recall of the details of the information presented by the robot than those to who do not contribute to the conversation (bystanders and overhearers).

Hypothesis 3 – Subjects whose presence the robot acknowledges and to whom it assigns through gaze cues a participant role (either as addressee or bystander) will evaluate the robot more positively than those whose presence the robot does not acknowledge and to whom it does not communicate a participant role (overhearers).

Hypothesis 4 – Subjects who contribute to the conversation as active participants (addressees) will express stronger feelings of groupness (with the robot and the other subject) than those who are not active participants of the conversation (bystanders and overhearers).

4.3.3. Experimental Procedure

Subjects were first given a brief description of the purpose and the procedure of the experiment. After the introduction, an experimenter asked them to review and sign a consent form. They were then provided with more detail on the task and asked to answer a pre-experiment questionnaire. Both subjects were told that researchers were developing a travel agent robot and would like their help in evaluating their design. Subjects were provided with identical instructions and randomly assigned to the conditions in the experiment. They were told that, after their interaction with the robot, they would be asked to answer a questionnaire on their experience and their recall of the material presented by the robot. After completing the task, subjects answered a post-experiment questionnaire that measured their recall of the information presented by the robot, their affective state, their perceptions of the robot, the group, and the task, and their demographic information.

The conversation with the robot and the experiment procedure in total took an average of 7.5 minutes and 25 minutes respectively. The experiment was run in a dedicated space with no outside distraction. A male native-Japanese-speaking experimenter was present in the room during the experiment. All subjects were paid ¥1,500 (approximately $14) for their participation.

4.3.4. Measurement

The manipulation in the robot's gaze behavior was the only independent variable. There were three types of dependent variables: behavioral, objective, and subjective.

4.3.4.1. Behavioral Measures

Subjects' conversational behavior was captured using high-definition cameras at 1080i resolution and stereo speakers. The video and audio data was coded for two behavioral measures of conversational participation: whether subjects took turns to respond to the robot and how long they spoke.

4.3.4.2. Objective Measures

Subjects' recall of the information presented by the robot was measured using a post-experiment questionnaire. The questionnaire included factual statements about information about the travel destination that the robot presented to the participants. Participants were asked to rate these statements as "true" or "false."

4.3.4.3. Subjective Measures

Subjective measures captured subjects' affective state using the PANAS scale (Watson et al., 1988), perceptions of the robot's physical, social, and intellectual characteristics using a scale developed to evaluate humanlike agents (Parise et al., 1998), feelings of closeness to the robot (Aron et al., 1992), feelings of groupness and ostracism (Williams et al., 2000), perceptions of the task (e.g., how much they enjoyed and attended to the task), and demographic information.

The subjective evaluation also included a question for manipulation check; subjects rated how much they thought the robot looked towards them and towards the other subject. Additionally, single-item measures assessed how much subjects thought the robot ignored them and considered their preferences in providing travel information. Seven-point Likert scales were used in all questionnaire items.

4.3.4.4. Participant Sample

Research in nonverbal behavior reports strong effects of group composition on both the production and the perception of gaze, particularly of gender (Exline, 1963; Argyle & Dean, 1965; Argyle & Ingham, 1972) and age (Efran, 1968; Libby, 1970). The first study of this dissertation (described in Chapter 3) also found gender effects on how the robot's gaze affected people's performances and their perceptions of the robot. One of the limitations of the first study was that it used observations of a female speaker in an all-female triad to design the gaze behavior of the robot and evaluated the designed gaze behavior with a mixed-gender population. I speculated that gender-based differences in the production and perception of gaze behavior would have an effect on the results of this study. Therefore, the current study intended to control for these group composition effects and test the hypotheses in a smaller population, particularly male college students from the Osaka area of Japan between the ages of 18 and 24. Accordingly, the subject profile for the observation was also limited to an all-male triad (and a male experimenter administered the study).

A total of 72 subjects participated in the experiment in 36 trials. All subjects were native-Japanese-speaking university students recruited from the Osaka area. The ages of the subjects varied between 18 and 24 with an average of 20.8 years. Subjects were chosen to represent a variety of university majors. Of all the subjects, 26 studied management sciences, 23 studied social sciences & humanities, 16 studied engineering, 5 studied natural sciences, and 2 did not report their academic majors. Subjects were randomly assigned to the experimental conditions. The computer use among subjects was very high (M=6.27, SD=0.98) on a scale from 1 to 7. Their familiarity with robots was relatively low (M=2.97, SD=1.67), so was their video gaming experience (M=2.92, SD=1.91). Five (out of 72) subjects had toy robots and 23 owned pets.

4.3.5. Results

Behavioral, objective, and subjective measures were analyzed using an analysis of covariance (ANCOVA). This method, similar to analysis of variance (ANOVA), applies a linear regression on the dependent variables that are significant across conditions to identify the direction of main effects and interactions while taking covariates into consideration that might account for some of the variance in data. This method was chosen to account for possible interactions between the two subjects in each trial. For instance, the number of speaking turns taken by one of the subjects is affected by the number of turns taken by the other subject in the same trial given that the robot yielded a fixed number of turns. In this situation, the analysis of covariance compared the number of turns taken by subjects with different participant roles while accounting for the number of turns taken by other subject in the same trial. From the statistical modeling point of view, for each dependent variable, data from subjects with different participant roles (overhearers, bystanders, and addressees) were entered into the model as response variables and data from the other subject (addressees) were entered in the model as covariates. In the third condition, because both subjects were addressees, data was randomly sampled into response variables and covariates in equal size. In the figures hereafter, the response variables are indicated with vertical stripes, horizontal stripes, and diagonal stripes for overhearers, bystanders, and addressees respectively. Covariates are indicated with no texture. An ID number for each pair of subjects was also included in the model as a random effect. Item reliabilities for scales and correlations across dependent measures were also calculated. Below, results of the analyses of each set of measures are provided.

4.3.5.1. Behavioral Measurements

The analysis of the behavioral data first looked at whether subjects to whom the robot yielded speaking turns took these turns. This analysis showed that subjects correctly interpreted these signals 98.71% of the time (307 of 311 turn-yielding signals) and conformed to them by taking speaking turns 97.11% of the time (302 of 311 turns). Of the nine turn-yielding signals to which they did not conform, six were passed between subjects (some addressees passed their turns to overhearers because they felt awkward talking to the robot while other subject was being ignored), three were not taken by the subjects due to ambiguities in robot's speech (in three trials, subjects did not perceive one of the questions as a question), and two were taken by both subjects as surprised responses to information presented by the robot (e.g., "Oh, I didn't know

	Condition 1 1 Overhearer 1 Addressee **Mean** (StDev)	Condition 2 1 Bystander 1 Addressee **Mean** (StDev)	Condition 3 2 Addressees **Mean** (StDev)
Number of speaker turns (counts)	**0.33** (1.15) **7.50** (1.17)	**1.08** (0.29) **7.75** (0.45)	**4.54** (1.82)
Total time spent speaking (seconds)	**0.60** (2.09) **9.43** (2.17)	**1.38** (0.66) **10.00** (3.19)	**6.09** (3.48)

Table 4.2. The number of turns participants took (top row) and the total time they spoke (bottom row) in each participant role in each condition.

that.")—these responses did not seem to be attempts to take the floor. Table 4.2 summarizes mean and standard deviation values for the number of speaking turns that subjects took and the total time they spent speaking for each participant role in each condition. The non-zero values for the overhearers in both measures are due to the six turns that addressees passed to them. Bystanders took an average of one turn as they responded to the robot during greetings.

Next, the analysis compared participation behavior across the three conversational roles by applying an analysis of covariance on the number of speaking turns that subjects took and the total time they spent speaking across the three conditions. Pairwise comparisons fully supported the first hypothesis. Addressees took significantly more speaking turns ($F[1,30]=17.58$, $p<0.01$) and spoke significantly longer ($F[1,30]=7.41$, $p=0.01$) than bystanders and overhearers. They also took significantly more speaking turns ($F[1,30]=6.75$, $p=0.01$) and spoke significantly longer ($F[1,30]=5.11$, $p=0.03$) than bystanders alone. No significant differences were found between bystanders and overhearers. Figure 4.11 illustrates these comparisons.

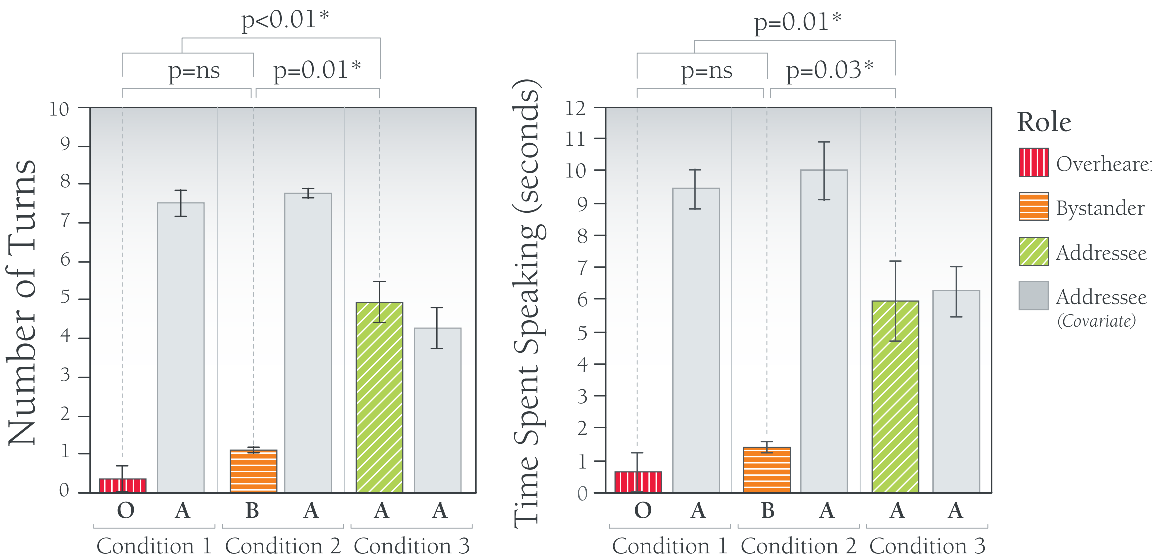


Figure 4.11. The number of turns subjects took (left) and the total time they spent speaking (right) in each conversational role.

4.3.5.2. *Objective Measurements*

The second hypothesis predicted that addressees would have better recall of the information presented by the robot than bystanders and overhearers. This prediction was not supported by the analysis; there were no significant differences across conditions in how well subjects recalled the information presented by the robot. The numbers of correct answers out of eight questions regarding the information that the robot presented on average were 2.75 (SD=1.66), 3.83 (SD=1.59), and 3.17 (SD=1.47) for overhearers, bystanders, and addressees respectively. While participant role did not affect subjects' recall of information, it affected their ratings of how much they attended to the task. Addressees rated themselves as attending to the conversation significantly more than bystanders and overhearers did, $F(1,29)=12.90$, $p<0.01$. These results are illustrated in Figure 4.12. Furthermore, the analysis found a strong effect of the topic of conversation (the travel destination) on recall of information, $F(1,33)$=

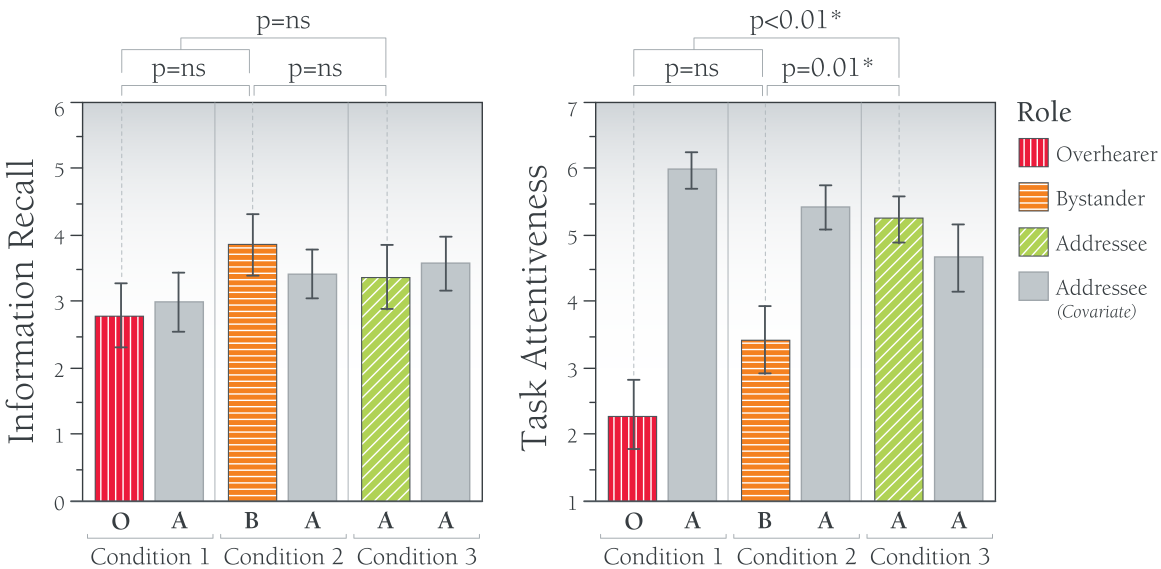


Figure 4.12. Subject's information recall (left) and task attentiveness (right) in each conversational role.

10.67, $p<0.01$. The effect of participant role on attentiveness to the task and the effect of travel destination on information recall provide some insight into why the third hypothesis was not supported by the results, which is further considered in the Discussion section of this chapter.

4.3.5.3. Subjective Measurements

The analysis of the data from subjective measures first tested whether the gaze manipulation was successful through a manipulation check, which was calculated by taking the difference between subjects' ratings of how much the robot looked at them and their ratings of how much it looked at the other subject. Pairwise tests compared these ratings between pairs of different participant roles across and within conditions. A successful manipulation would mean that there would be no differences between the ratings of the two addressees in the third condition and significant differences in all other pairwise comparisons. The results of the analysis supported these

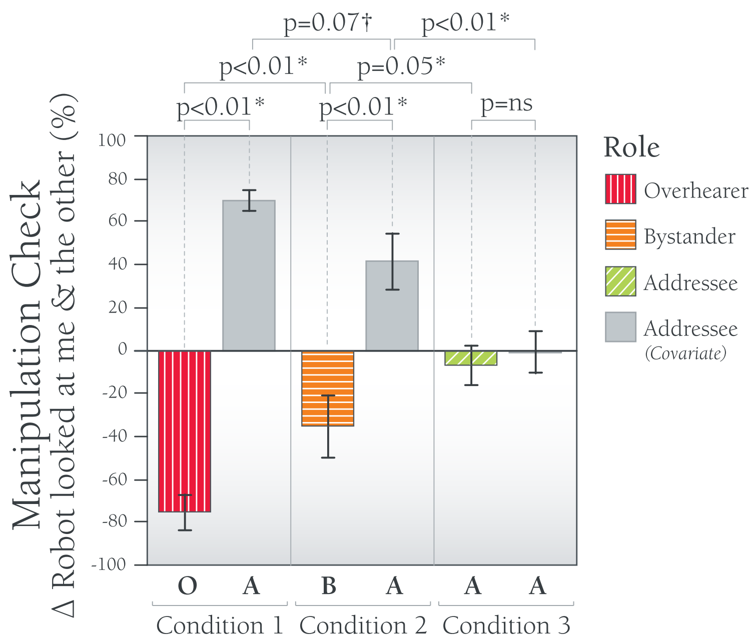


Figure 4.13. Manipulation check measured by the difference between how much participants thought that the robot looked at them and how much they thought it looked at the other subject.

predictions. No differences were observed between the addressees in the third condition and all other comparisons were statistically significant with a marginal difference between ratings of bystanders and overhearers. Figure 4.13 provides results for all pairwise tests.

Next, item reliabilities were calculated for the two main measures that tested the third and fourth hypotheses. Item reliabilities for the three-item scale that measured how much subjects liked the robot (Cronbach's α=0.76) and the six-item scale for measuring feelings of groupness (Cronbach's α=0.92) were sufficiently high.

The third hypothesis predicted that subjects whose presence the robot acknowledges (addressees and bystanders) would like the robot more than those whose presence it does not acknowledge (overhearers). An analysis of covariance on subjects' liking of the robot supported for this prediction. Addressees and bystanders liked the robot significantly more than overhearers did, $F(1,30)=7.35$, $p=0.01$. Bystanders also liked the robot significantly more than overhearers did ($F[1,30]=4.05$, $p=0.05$), suggesting that the simple acknowledging gaze led subjects to like the robot more. There were no significant differences in addressees' and bystanders' liking of the robot. Figure 4.13 illustrates the results from these comparisons.

The fourth hypothesis was also supported by the analysis. As predicted, those who were placed in the role of addressee by the robot and who contributed in the conversation as active participants rated their feelings of groupness significantly higher than those who did not contribute to the conversation as bystanders (except during greetings and leave-taking) or as overhearers, $F(1,30)=8.95$, $p<0.01$. Addressees also rated their feelings of groupness as higher than bystanders alone ($F[1,30]=5.36$, $p=0.03$) and overhearers alone ($F[1,30]=8.25$, $p<0.01$). These comparisons are also illustrated in Figure 4.14.

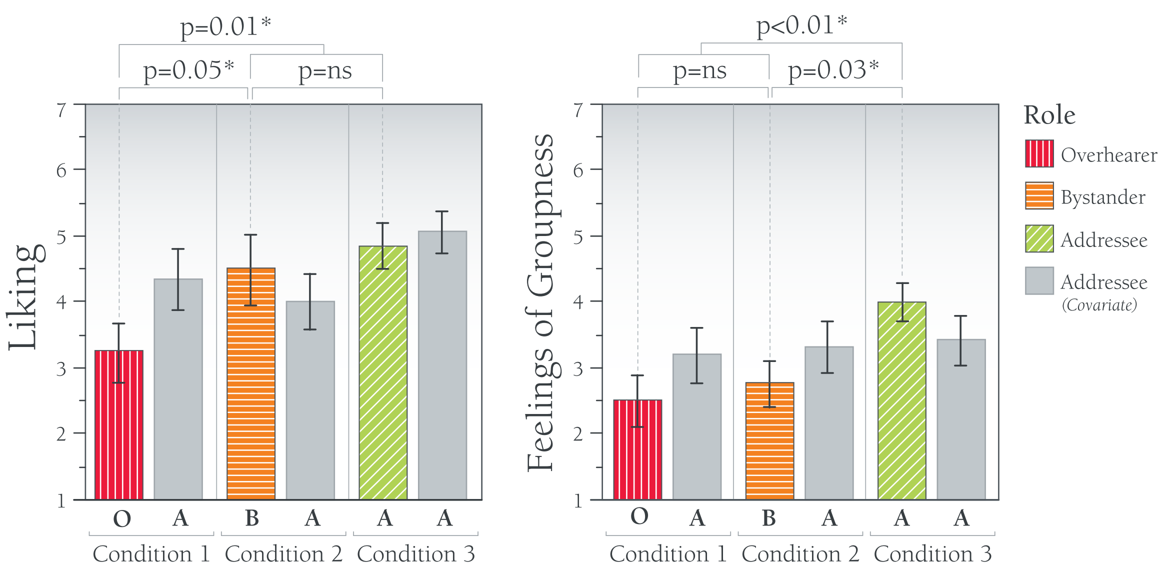


Figure 4.14. Subject's information recall (left) and task attentiveness (right) in each conversational role.

The analysis of the data from single-items scales (on how much subjects thought the robot ignored them and considered their preferences in providing travel information) provides further explanation for why overhearers liked the robot less than others did and why addressees felt more feelings of groupness than others did. Subjects whom the robot ignored did, in fact, feel significantly more ignored than both bystanders ($F[1,30]=4.41$, $p=0.04$) and addressees ($F[1,30]=14.14$, $p<0.01$) did, which perhaps led to their liking the robot less. Similarly, addressees, who contributed to the conversation more than others did, thought that the robot considered their preferences significantly more than bystanders ($F[1,30]=4.05$, $p=0.05$) and overhearers ($F[1,30]=6.98$, $p=0.01$) did. This mutual exchange conceivably led to more cohesion in the group as reflected in subjects' feelings of groupness.

Finally, Pearson product-moment correlations were calculated to understand how dependent variables related to each other. These analyses showed statistically significant correlations between familiarity with robots and liking (r=0.26, p=0.03), task attentiveness ($r=0.25$, $p=0.04$), and feelings of groupness, $r=0.37$, p<0.01.

4.3.5.4. Qualitative Observations

Qualitative observations of subjects' interactions with the robot shed further light on the quantitative results. In these observations, subjects did not speak unless they were granted a turn, with the exception that in three of the trials addressees showed in their nonverbal behavior hesitation and discomfort with the robot ignoring the other conversational partner. This could be seen in repeated glances toward the ignored subject, perhaps to see the reaction to this unfair treatment. They alleviated this discomfort by passing some of their speaking turns to overhearers. While this behavior is a breakdown in the participant structure established by the robot, I reason that it also illustrates how well people conformed to the signals that the robot communicated to them. Subjects who were not passed up on speaking turns by the robot still did not attempt to take turns unless they were passed up by the other subject. Similarly, those to whom the robot yielded turns knew that they had the floor and felt the liberty to pass their turns to the other subject.

In a number of trials, subjects hesitated to take the speaking turn after they received the first turn-yielding signal from the robot. One explanation of this behavior is that the subjects were not sure that the robot could understand them. Another explanation is that they felt uncomfortable talking to a robot in front of the experimenter and the other subject. This behavior was not observed after the first turn exchanges in these trials, perhaps because they were assured with their experience with the first turn exchange that the robot could understand them.

When responding to the robot, people often used articulate language—full sentences instead of phrases. They also produced gaze signals similar to those observed in human communication. For instance, human communication research has found that "breaking mutual gaze" (looking away from the speaker) when answering questions is a common behavior (Libby, 1970). In the experiment, subjects broke mutual gaze with the robot before replying to 35.37% of the robot's questions and 47.12% of the questions that required them to make an evaluation (e.g., choosing a travel destination) before answering. This behavior provides some evidence that the subjects perceived the turn-yielding gaze cues from the robot as valid social stimuli and responded to these signals by creating the appropriate communicative behavior.

4.4. Discussion

Drawing on theory on conversational organization and formal observations of human conversations, I modeled three kinds conversation-structural signals that speakers communicate by means of gaze: information structure (thematic organization), conversation structure (turn-exchanges), and footing structure (participant roles). I recreated these signals in a humanlike robot as a part of a fluid conversional behavior and contextualized this behavior in a human-robot conversation scenario.

The experimental evaluation supported three of my four hypotheses. Table 4.3 provides a summary of these predictions and whether they were supported by the results. Using only gaze cues, the robot manipulated who participated in and attended to a conversation, subjects' feelings of groupness, and their liking of the robot. Subjects accurately read the robot's turn-yielding gaze signals 99% of the time and conformed to these signals by taking 97% of the speaking turns. People also conformed to the participant roles that the robot communicated to them. Those whom the robot treated as addressees took more speaking turns and spoke longer than those who were treated as bystanders or as overhearers. Addressees also attended to the task more and felt stronger feelings of groupness than others. Those whose presences were acknowledged as addressees or as bystanders liked the robot more than those who were ignored as overhearers. Contrary to my prediction, participant role did not affect subjects' recall of the information by the robot.

Hypotheses	**Results**
Subjects' compliance with the robot's footing signals	
Hypothesis 1 – Subjects will correctly interpret the footing signals that the robot communicates to them and conform to these roles in their participation to the conversation.	Supported
The effect of participation in the conversation on information recall	
Hypothesis 2 – Subjects who contribute to the conversation by taking speaking turns (addressees) will recall the details of the information presented by the robot better than those who do not contribute to the conversation (bystanders and overhearers).	Not supported
The effect of participant role on liking of the robot	
Hypothesis 3 – Subjects whose presence the robot acknowledges and to whom it assigns a participant role (either as addressee or bystander) will evaluate the robot more positively than those whose presence the robot does not acknowledge and to whom it does not assign a participant role (overhearers).	Supported
The effect of participant role on feelings of groupness	
Hypothesis 4 – Subjects who contribute to the conversation as active participants (addressees) will express stronger feelings of groupness (with the robot and the other subject) than those who are not active participants of the conversation (bystanders and overhearers).	Supported

Table 4.3. Summary of hypotheses and whether they were supported by the results.

Further analyses of the objective and subjective measures provide some insight into why the prediction on information recall was not confirmed. The analyses show that addressees rated their attentiveness to the task higher than others did. While it is conceivable that attentiveness should lead to better recall of information, the finding that the topic of the conversation significantly affected information recall suggests that subjects' prior knowledge of the topic might have been too well established to be affected by the information presented by the robot. Administering a pre-experiment questionnaire to measure prior knowledge of the topic would have helped in identifying how much new information was learned during the experiment. Alternatively, choosing a conversation topic, such as a fictional story, on which subjects would have sparser pre-existing knowledge could have provided greater support for my predictions.

Some of the findings from the analyses of human gaze data can be generalized to the design of other conversational systems. The three structures that were identified in the analysis (information, conversation, and footing structures) will affect speaker's gaze shifts differently based on the participation structure of the conversation. In an oratorial setting, such as a lecture or storytelling, information structure or thematic organization will account for the majority of gaze shifts, as in the first study of the dissertation (Chapter 3). In a two-party conversation, conversation structure (turn-exchanges) will also be an important element of the design of the speaker's gaze shifts. Footing structure (participant roles) will affect speaker gaze in conversations with two or more speakers based on the roles of interlocutors. For instance, in a two-party conversation in which the participants' footings are equal, the design of the gaze behavior may not have to account for participation structure. On the other hand, in a two-party conversation in which one of the participants holds the floor for extended periods or in a conversation with multiple parties with different participant roles, participation structure will be an important part of the design of the speaker's gaze behavior.

4.4.1. Limitations

The results presented here have a number of limitations. First, because only male subjects were used, the results have limited generalization to conversational situations with female subjects or mixed-gender groups. Ideally, a gender-balanced, full-factorial-design study is required to understand how gender affects participation structure in human-robot conversations. Secondly, these results might not generalize beyond the cultural context of the study. Factors such as Japanese subjects' possible closer familiarity with robots and the frequent use of interfaces that use speech in Japan might have affected the results. In fact, contrary to the results of this study, the first study of the dissertation (conducted with a American population) showed that people's liking of the robot was significantly correlated with video gaming experience and not with familiarity with robots, suggesting fundamental differences in how people might perceive and interact with robots across the American and Japanese cultures. Furthermore, differences in conversational conventions—particularly those brought about by age, social status, organizational rank, and so on—across cultures might affect these results. Our understanding of these cultural differences would greatly benefit from cross-cultural studies of human-robot interaction. These cross-cultural limitations and implications are further considered in the General Discussion chapter.

The generalizability of these results also suffers from the limited interactivity of the robot, which forced the design of the conversational scenario to have the robot hold the floor for most of the conversation and yield turns only at scripted points in the conversation. The results of this study might have been different with a more fluent conversational scenario where participants took more turns and held the floor for longer periods. Robust speech recognition and adaptive speech generation would allow for exploration of unscripted, fluent conversational scenarios.

Because subjects were not told that they might be assigned different participant roles, they might have felt the need to further regulate the roles that the robot communicated to them. In three trials, addressees passed some of their turns to overhearers. I argue that these subjects expected to be treated as equals by the robot—subjects' equal body orientations relative to the robot further supported this expectation—and the robot's ignoring one of the subjects caused some discomfort. They might have tried to alleviate this discomfort through passing some of their speaker turns to the ignored subject. While this behavior shows the effectiveness of the robot's gaze behavior in signaling who is granted the next turn, it also highlights the ever-changing nature of participant roles in conversations as also emphasized by Goffman (1979). This behavior also shows the importance of context in adapting participant roles. It was important for this study that subjects were given minimal information about the nature of the study as I wanted to test how well the robot could communicate to subjects their participant roles. I argue that the dynamic nature of participant roles and the role of context pose fruitful areas for future research on human-robot conversations.

Another set of limitations is imposed by the methodology. Firstly, the design of the robot's gaze behavior was informed by data collected from a single speaker. While this choice allows for an in-depth analysis of the data, it also imposes a limitation on the generalizability of the results of these analyses. Secondly, the designed speaker gaze behavior did not consider the behaviors of the addressees. This decision imposes a significant limitation on the designed behavior and the results of the evaluation. Accounting for addressee behavior in directing the robot's gaze behavior in the experiment required instrumenting participants with eye-trackers. However, the use of eye-trackers may have introduced an additional source of error and exasperated the awkwardness of conversing with a robot. Future exploration of how addressee behavior might affect speaker gaze behavior would benefit from robust camera-based eye-tracking systems.

While one of the goals of the theoretically and empirically grounded approach is formalizing the design process, the designed behaviors are still significantly influenced by the decisions made in the analyses. For instance, my choice of unitizing the speech data at points of thematic transition—creating a rather large unit of analysis—forced me to seek patterns initiated by the onset of a "thematic field" in the speaker's gaze behavior. A smaller unit of analysis (such as intonation units that represent the prosodic structure of speech) could have led to closer coupling between information structure and gaze shifts in the designed gaze behavior. This sensitivity to differences in designer's choices in studying empirical data can be addressed by verifying the outcome of these design decisions through intermediary user studies. The design process would significantly benefit from intermediary evaluation stages. This limitation is also further considered in the General Discussion chapter.

4.5. Study Conclusions

During conversations, people use gaze cues to establish and maintain their and their conversational partners' participant roles or "footing." This study showed how these cues can be used by a robot to regulate footing in human-robot conversations. Using findings from human communication theory and formal observations of human behavior, I designed gaze behaviors for a robot to cue three kinds of participant roles: addressee, bystander, and overhearer. A controlled laboratory experiment conducted with 72 subjects in 36 trials showed that these cues affected subjects' participation in a conversation with the robot, how much they attended to the conversation, how much they liked the robot, and how strongly they felt a part of the group that included the robot and their conversational partners.

Behavioral measures showed that subjects correctly interpreted 99% of the turn-yielding signals and took 97% of these turns. Those who took turns as active participants of the conversation rated their attentiveness to the conversation higher than those who did not take speaking turns did. They also felt more acknowledged, welcomed, and valued by their group and that they belonged more to the group than those who remained as non-participant bystanders or overhearers. Bystanders, whose presence the robot acknowledged with simple non-turn-yielding gaze signals, evaluated the robot more positively than overhearers, for whom the robot did not produce these signals.

While results of this study are limited to the cultural and conversational context of the study and the characteristics of the studied population, they do provide evidence on how robots might use gaze cues for shaping participant roles in conversations. Further work is required to generalize these results and gain a fuller understanding of how gaze cues relate to conversational organization in human-robot interaction.

The next chapter describes the third and last study of the dissertation, which looks at designing mental state-communicating gaze cues.

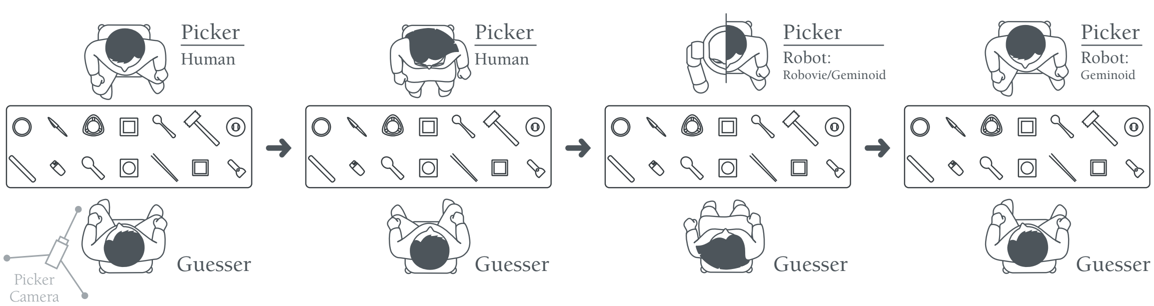
5. Study III: Designing Gaze Cues to Communicate Mental States

This study explores how robots might use gaze cues to communicate mental states, an important function that gaze serves in human communication. Understanding how robots might perform this function using gaze cues could provide a number of social and cognitive benefits. First, to develop the ability to read the mental states of others through their social cues is an important developmental and educational goal. As envisioned in the scenario where Ken played educational games with a robot designed to help develop the ability to read subtle social cues to interpret mental and emotional states (Section 1.1.3), this research could inform the design of robotic applications to facilitate these developmental goals. Second, gaze cues that communicate mental states are still not well understood in human communication, mainly due to the lack of experimental methods that allow for a precise control of gaze stimuli. This research could harness a robot's ability to produce precise, controllable social stimuli and contribute to the understanding of the relationship between gaze cues and attributions of mental states. Finally, this exploration could introduce a rich design space for creating natural humanlike social behaviors.

To move toward these goals, this study focuses on a particular family of gaze cues called "nonverbal leakage," which includes a set of non-strategic, non-semantic cues that are produced unintentionally and that reveal the mental states of an individual. Because these cues are still not well understood in human communication and the theoretical and methodological foundations for their study is not well established, this study first tries to gain a better understanding of how humans produce these cues and the social and cognitive outcomes of their production using an experimental framework developed specifically for this study. In this framework, members of dyad play a guessing game, one playing as the picker and the other playing as the guesser. The picker mentally picks one of the items placed on a table. The guesser then asks the picker a series of questions that can be answered with "yes" and "no." Once the guesser collects enough information about the item, he or she guesses which item the picker had picked. The design of this experimental framework provided the setup to study whether the picker, under cognitive pressure to correctly answer the questions while not revealing the pick, would "leak" information through gaze cues.

This chapter starts with a description of this experimental framework and a summary of the results from a short study conducted to model how pickers leaked information through their gaze (Section 5.1). The modeling found that pickers frequently produced very short glances at their picks before answering questions. This short modeling step is followed by the first experiment of the study that investigated whether pickers' production of leakage cues would lead guessers to guess the item faster by manipulating whether guessers could see the pickers' eyes and measuring performance effects (Section 5.2). The results showed that guessers identified the item using fewer questions when the pickers' eyes were visible. The second experiment studied whether a robot's production of these cues would lead guessers to guess the robot's pick faster and measured how the design of the robot might affect the results by comparing the effects of two robots with different levels of humanlikeness (Section

5.3). Participants identified the robots' picks faster and used fewer questions when the robots produced leakage cues. This result was significant only for Geminoid and not for Robovie. The last experiment was a follow-up of the second experiment, investigating whether a robot's production of leakage cues would affect guessers' subjective evaluations of the robot (in addition to their performance) and how these perceptions would be affected by whether the robot tried to conceal the leaked information (Section 5.4). The results showed that participants identified the item using fewer questions when the robot produced leakage cues but not when it produced concealing cues. Concealing cues also led to ratings of less cooperativeness. The last section summarizes the findings from all three experiments and discusses the



Modeling (Section 5.1)	Experiment I (Section 5.2)	Experiment II (Section 5.3)	Experiment III (Section 5.4)
Human-human interaction	Human-human interaction	Human-robot interaction	Human-robot interaction
2 Japanese dyads	12 American dyads	26 Japanese participants	48 Japanese participants
No manipulation	**No gaze cues** vs. **gaze cues**, between participants	**No gaze cues** vs. **leakage cues**, within participants; **Geminoid** vs. **Robovie**, between participants	**No gaze cues** vs. **leakage cues** v.s. **concealing cues**, between participants
Leakage cue: Pickers shortly glanced at their picks before answering guessers' questions. **Concealing cue:** They subsequently glanced at other items to conceal.	Guessers identified the item using fewer questions when they could see picker's gaze cues.	Guessers identified the item faster when the robot leaked, but only with Geminoid.	Guessers identified the item using fewer questions when the robot leaked, but not when it concealed. Concealing cues led to lower ratings of cooperativeness.

Figure 5.1. A summary of the experiments involved in this study.

research implications of the study. Figure 5.1 provides a summary of the experiments involved in this study.

5.1. Theoretically and Empirically Grounded Design

Leakage cues are nonverbal acts that give away information about the mental and emotional states of an individual, particularly about internal states and intentions that the individual wishes to hide (Ekman & Friesen, 1969, 1974; Zuckerman et al., 1981). To gain a better understanding of how people might leak information through gaze cues, to design such behaviors for a robot, and to evaluate whether people can read these cues in robots and interpret them correctly, I devised an experimental paradigm in which a dyad—either two participants or a participant and a robot—played a game of guessing. In the game, one of the players, the "picker" mentally (without identifying it to the other player) chose an item from among fourteen items placed on a table located between the two players (see Figure 5.2). The other player,

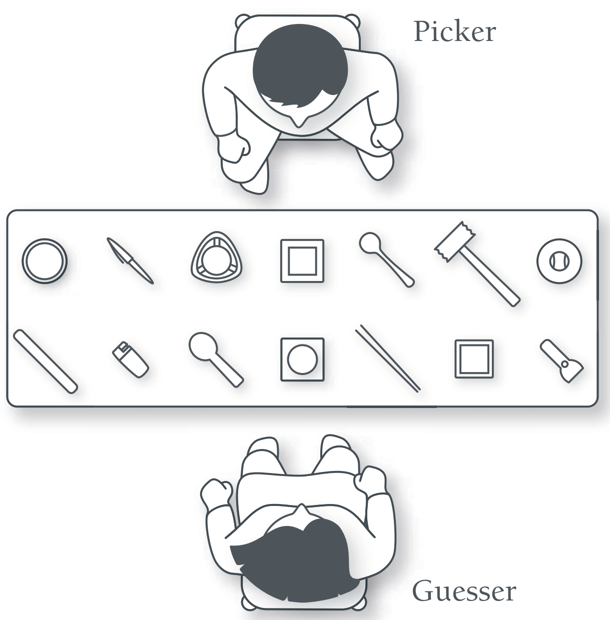


Figure 5.2. The setup of the guessing game.

the “guesser,” tried to guess which item the picker chose by asking the picker a set of questions that could be answered with “yes” or “no.”

The items on the table were carefully chosen from artifacts commonly used in daily life and that represent a balanced set of colors, shapes, materials, and sizes. These items were placed on the table equidistantly and their spread was determined so that the players did not have to move their heads to glance at the items. Participants were provided with detailed instructions and strategies on how to play the game. They were told that the best way to play the game was to ask questions that would help them narrow down the number of alternatives. For instance, if they asked whether the item has the color red and the picker said, “Yes,” this would reduce alternatives from fourteen to four; if the picker said, “No,” the number of alternatives would still be reduced to ten. The number of items on the table was empirically determined in a pretest to be fourteen to allow participants to identify the item with an expected average of five questions.

5.1.1. Leakage Gaze Cues

To understand whether—and if so, how—human pickers would “leak” information about the items that they chose, two all-male dyads were hired to play the game (see Figure 5.3).Participants played the roles of the picker and the guesser in eight sessions for each role. The pickers’ gaze behaviors were captured using high-definition cameras, and a frame-by-frame analysis of the video sequences was conducted.

The most significant finding of the analysis was that the pickers often—as frequently as a total of 22 times in an eight-session game—gazed in very short glances toward their pick immediately before answering the guessers’ questions. I speculate that the pickers glanced at their picks to verify whether the response to the guessers’ questions

Figure 5.3. Participants playing the game of guessing.

should be “yes” or “no” while trying to conceal their glance so as to not give away valuable information about their picks to the guessers. Ekman and Freisen (1969) describe two kinds of responses in the context of deception; when deceivers realize that their deception might be discovered, they either try to continue their deception, but “conceal” it, or “give up” their deception. These two behaviors were observed during these points of leakage in the data:

Leakage gaze cue – Pickers glanced at their picks for a very short duration immediately before answering the guessers ‘questions or, less commonly, immediately after answering the questions. Figure 5.4 shows image sequences of a participant producing a leakage gaze cue.

Concealing gaze cue – Pickers glanced at their pick for a very short duration immediately before answering the guessers’ questions followed by one or more

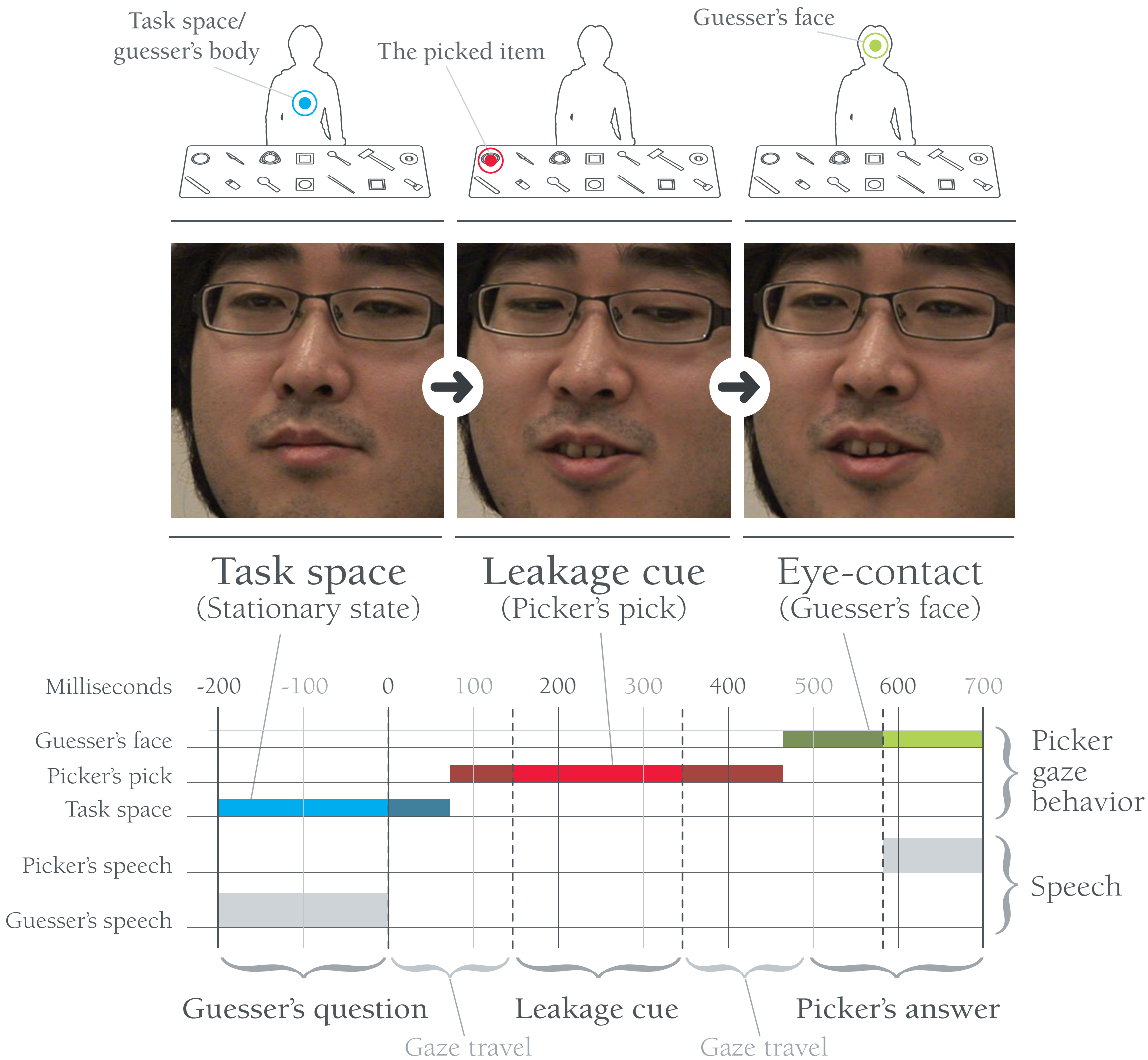


Figure 5.4. A participant producing a leakage gaze cue.

subsequent glances at other items on the table, concealing the pick. Figure 5.5 shows image sequences of a participant producing a concealing gaze cue.

5.1.2. The Design of Leakage and Concealing Gaze Cues for Robots

The results from the analysis of the human data directly informed the design of leakage gaze cues and concealing gaze cues for two robots: Robovie R-2 and

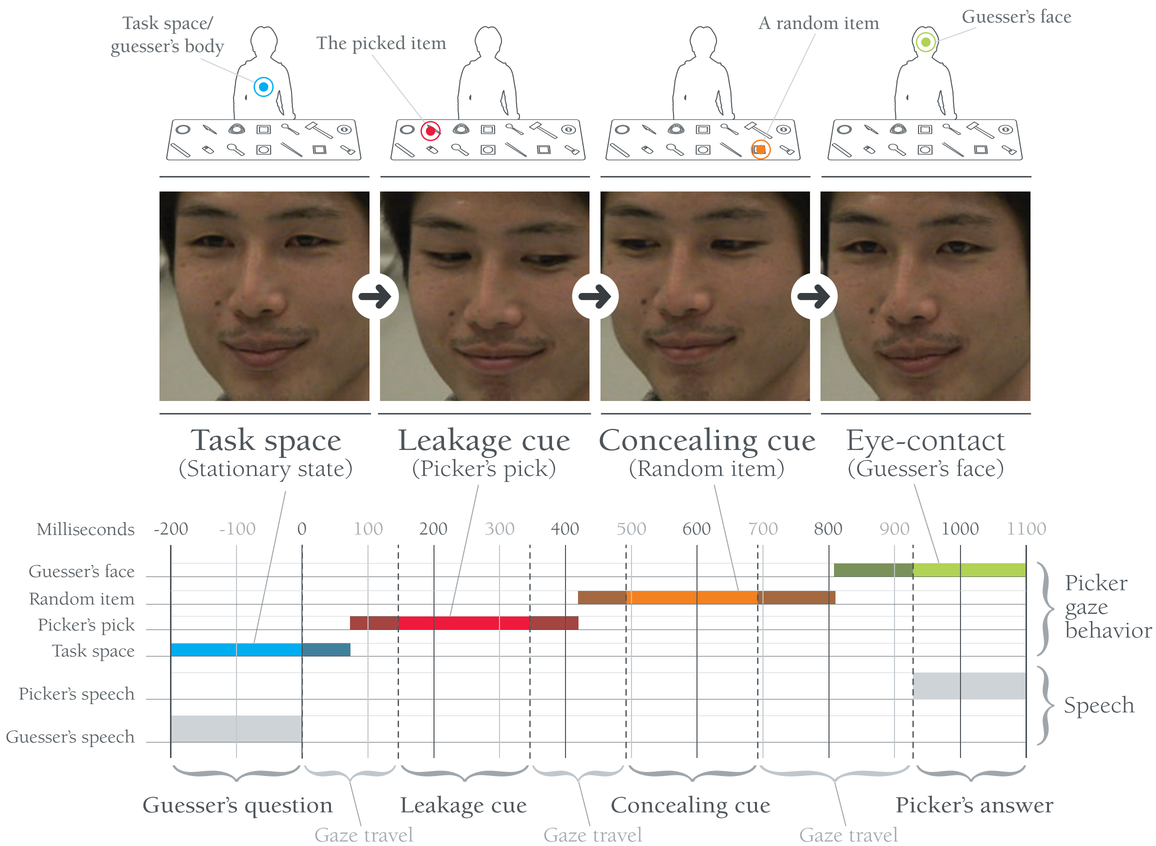


Figure 5.5. A participant producing a concealing gaze cue.

Geminoid, a highly humanlike android (Nishio et al., 2007). Two cues were designed for the robots:

Leakage gaze cue – The robots glanced toward their picks immediately before answering two of the first three questions that their partners directed to them. Figure 5.6 shows image sequences of the two robots producing leakage gaze cues.

Concealing gaze cue – The robots produced two subsequent glances, one toward their pick and one toward another item on the table, immediately before answering two of the first three questions that their partners asked. Figure 5.7 shows image sequences of the two robots producing concealing gaze cues.

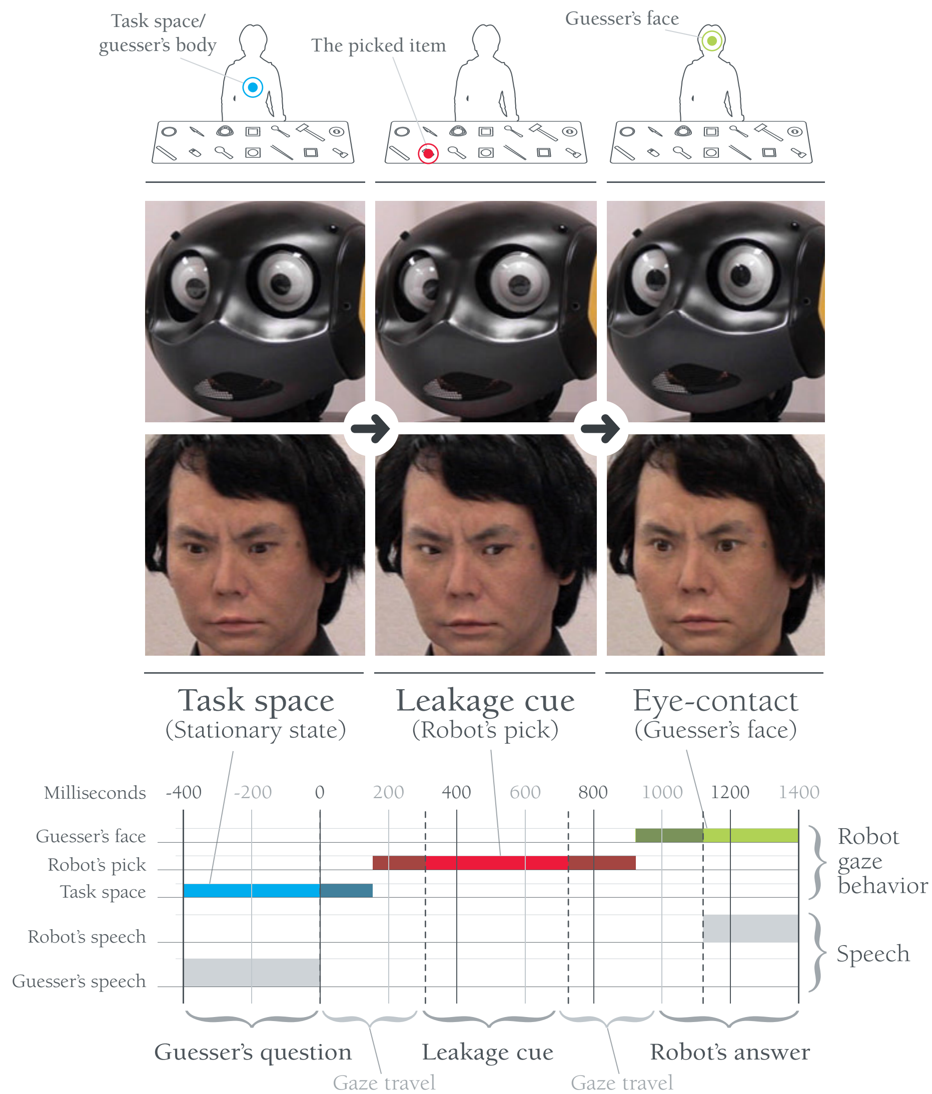


Figure 5.6. Robovie (top) and Geminoid (bottom) producing leakage gaze cues.

The lengths of the two kinds of glances were calculated by taking the gaze length distribution parameters from the human data as a basis and modifying these parameters to optimize for the motor capabilities of the two robots for smooth and natural motion and keep the total gaze durations for the two robots equal. Gaze length distribution parameters for leakage and concealing gaze cues calculated from human data and those created for the robots are provided in Appendix F.

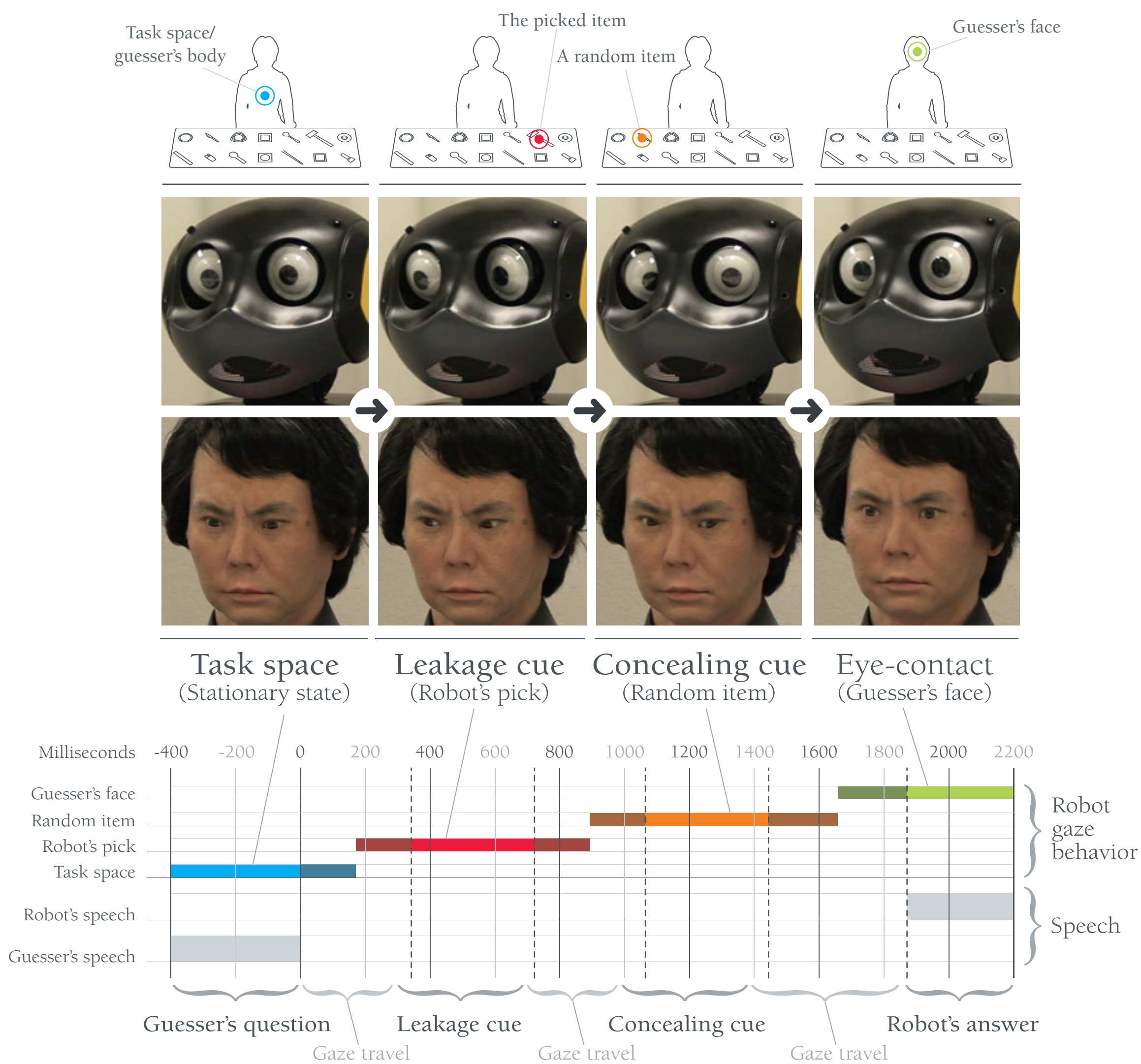


Figure 5.7. Robovie (top) and Geminoid (bottom) producing concealing gaze cues.

During their interaction with their partners, the robots followed common interaction rituals (Goffman, 1971; Forgas, 1979). They introduced themselves to their partners, provided them with information about the task, maintained fluency in the interaction by using phrases such as “Let’s play one more time,” and ended the interaction by thanking their partners for playing the game. The robots used a rich library of utterances, which was created by a human performer recording each expression several times in different forms and inflections. The performer’s lip movements were also captured using a five-camera motion-capture system to synchronize Geminoid's lip movements with its speech. In performing the task, the robots did not use speech recognition. Instead, a human operator initiated the robots' speech by selecting expressions from the library of utterances created for the robots.

The behaviors of the two robots were designed to be identical and follow the same pre-scripted routine and adaptive dialog, except for differences required by the physical design of the robot. One of these differences was that Geminoid produced eye blinks at an average interval of five seconds. Geminoid's voice was also differentiated in pitch from Robovie's to match the appearance of the robot, creating a low-pitch male voice for Geminoid and a high-pitch metallic voice for Robovie, which was done using post-processing in order to maintain the same length and inflections for each expression between the two robots.

5.2. Experiment I: Leakage Cues in Human-Human Interaction

The first experiment investigated whether people use information from their partners' gaze cues, including leakage gaze cues, to make inferences of mental states. It followed a two-by-one between-participants design in which two naïve participants played the guessing game and took turns playing the role of the picker and the guesser for 10 rounds following a practice round. The manipulated variable was whether the picker's gaze cues were visible to the guesser or hidden by reflective sunglasses, which blocked all leakage gaze cues. The gaze manipulation is shown in Figure 5.8.

5.2.1. Hypotheses

The goal of this experiment was to understand whether people use information from their partners' gaze cues to attribute mental states to them, which could be inferred

Condition 1:
Picker's gaze cues visible

Condition 2:
Picker's gaze cues not visible

Figure 5.8. Picker's gaze cues visible (left) and not visible (right) conditions.

from whether guessers show better task performance in the guessing game when these cues are available to them. I hypothesized that guessers would perform significantly better in identifying the pickers' picks—that is, use significantly fewer questions and/or take significantly less time to do so—when pickers' gaze cues are visible to them than when they are hidden from them.

5.2.2. Participant Sample

Twenty-four Carnegie Mellon University undergraduate and graduate students (14 males and 10 females) participated in the experiment. Participants' ages ranged from 19 to 28 (M=21.86, SD=2.15). Of the 24 participants, 9 majored in engineering and computer sciences, 8 studied social sciences and humanities, 4 majored in natural sciences, and 3 studied management sciences. Figure 5.9 shows participants playing the guessing game.

Figure 5.9. Participants playing the game under the condition in which gaze cues are not visible.

5.2.3. Experimental Procedure

The experimenter first provided the participants with information about the general purpose of the experiment—that the experimenters were designing a robot that could play games with people and would like to first observe how people play these games—and sought informed consent. Participants were then provided with the details on the game task. The experimental manipulation was introduced at this point; in half of the trials, participants were told that because the robot will not have eyes and the experimenters would like to best simulate this situation, they would be asked to wear reflective sunglasses. Participants then played a practice round of the game and asked any questions they might have had about the game rules or the characteristics of the items on the table. They then played 10 rounds of the game, switched roles with their partners, and played another 10 rounds. The game task took on average 10 minutes and 17 seconds (SD=3 minutes 23 seconds) and the complete experiment took an average of 20 minutes. Participants were paid $7 for their participation.

5.2.4. Measurement

The experiment involved a single manipulated independent variable: whether pickers' gaze cues were visible to the guessers. Two dependent variables measured *objective* task performance: (1) the number of questions guessers asked to identify the picked item and (2) the amount of time it took them to do so. All experimental trials were recorded with high-definition cameras. Video recordings were coded for how long game rounds took (starting with the picker's verbal signal that the picker picked an item and ending with the picker's confirmation that the guesser correctly guessed the item) and the number of questions that the guesser asked (those that were answered by the picker).

5.2.5. Results

The measurements of the dependent variables were analyzed using an analysis of covariance (ANCOVA). This method was chosen to model the interdependence between the two participants in each trial. For each dependent variable, the ANCOVA model used the data from each participant as the response variable and the data from the other participant in the same trial as the covariate. An ID number for each participant was also included in the model as a random effect. The task performance data included 245 trials. The distributions of the performance measures were transformed using the logarithm function to correct for the positive skew in the data distributions and outliers on the right tails of the distributions without losing data samples.

I hypothesized that guessers who played the game when their partners' gaze cues were visible to them would perform significantly better in identifying their partners' picks than those who played the game when their partners' gaze cues were not visible to them. The analysis of the number of questions measure confirmed this hypothesis. Guessers who could see their partners' gaze cues asked significantly fewer questions to identify their partners' pick than those who could not see their partners' gaze cues, $F(1,19)=6.11$, $p=0.02$ (Figure 5.10). This effect was not present in the time measure; the gaze manipulation did not have a significant effect on the time guessers took to identify their partners' picks, $F(1,15)=1.27$, p=ns.

The analysis did not show significant gender effects on the number of questions participants asked ($F[1,19]=2.25$, p=ns) and the time it took them to identify the pick ($F[1,19]=0.02$, p=ns). However, a main effect of gender configuration was identified on the number of questions asked ($F[2,18]=4.28$, $p=0.03$). Participants in female-female (FF), male-male (MM), and female-male (FM) dyads asked on average 4.36 (SD=1.18), 5.12 (SD=1.50), and 4.76 (SD=1.57) questions respectively. Furthermore, the visibility of the pickers' gaze cues significantly affected the number of questions

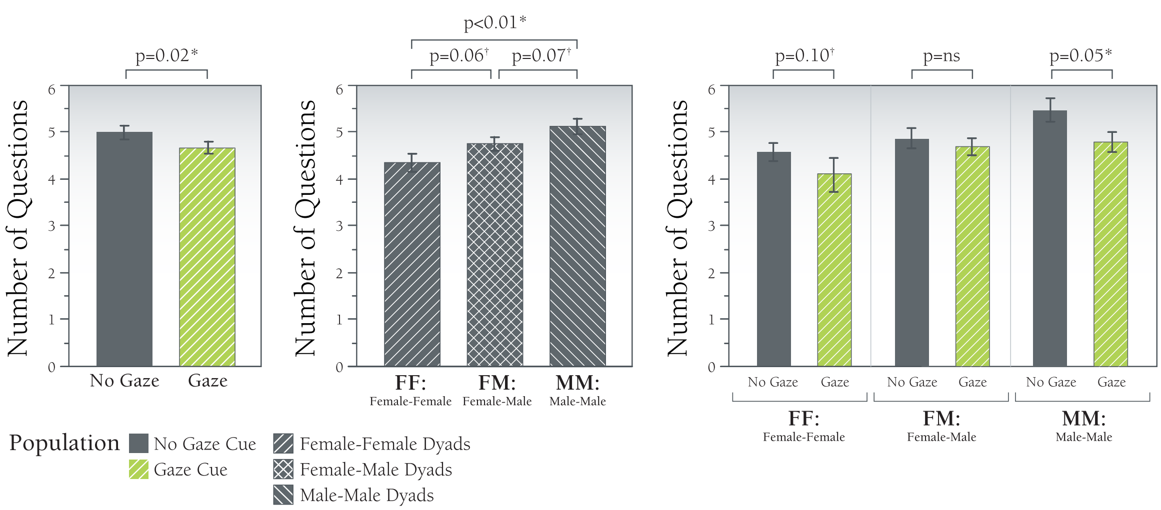


Figure 5.10. Number of questions asked in no gaze vs. gaze conditions (left), in FF, FM, MM dyads (middle), and FF, FM, MM dyads in gaze and no gaze conditions (right).

that participants asked in MM dyads ($F[1,18]=4.27$, $p=0.05$) and marginally did so in FF dyads ($F[1,20]=2.93$, $p=0.10$), while it did not affect the number of questions asked in MF dyads ($F[1,17]=0.21$, p=ns). Figure 5.10 shows pairwise comparisons between gaze and no gaze conditions in the different gender configurations.

5.2.6. Discussion

This experiment was aimed at understanding whether people use information from gaze cues and, therefore, tested the hypothesis that, in a game of guessing, guessers who could see their partners' gaze cues would perform better than those who could not see their partners' gaze cues. The results confirmed this hypothesis; guessers who played the game with partners whose eyes were visible to them correctly guessed their partners' pick using significantly fewer questions than those who played the game with partners who wore reflective sunglasses. This result suggests that people use information from others' gaze cues—including leakage gaze cues—to make inferences on their mental states.

The results also showed that gender combination significantly affected participant performance. Performances of participants in FF, FM, and MM dyads were as follows (from best to worst): FF > FM > MM. This ordering is parallel to the amount of mutual gaze that individuals show during conversations in these groups, as reported by Argyle and Ingham (1972): FF (38%) > FM (31.5%) > MM (23%). This alignment suggests an increased performance and mental state attribution with increased levels of mutual gaze, further supporting the argument that gaze cues are essential in communicating mental states. Furthermore, the results showed that the absence of gaze cues significantly affected only the performances of individuals in the MM dyads.

This result is consistent with gaze research; people rely on gaze cues more when gaze levels are particularly low (Beattie, 1980).

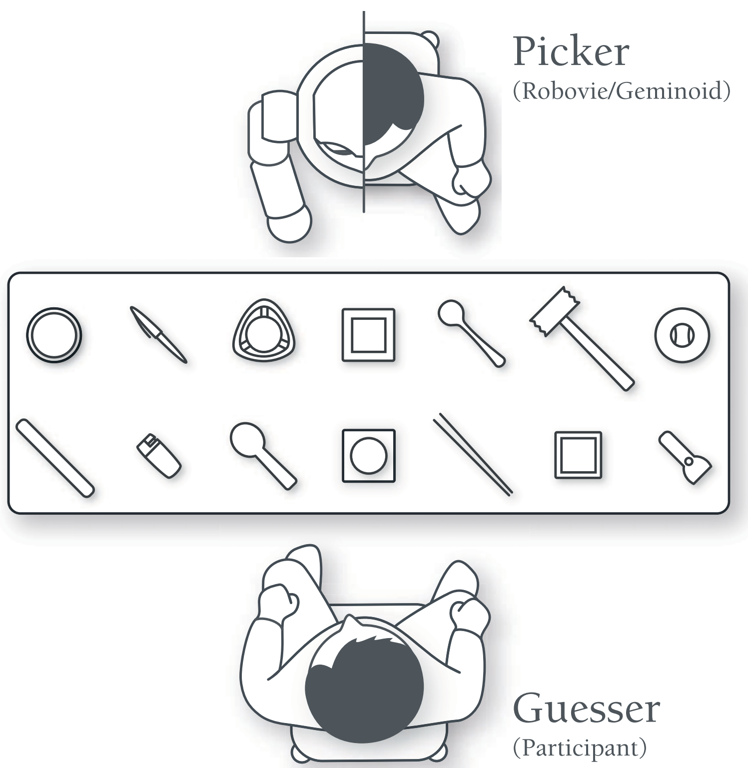
5.2.6.1. Limitations

A factor that might have contributed to participants' performing better when their partners' eyes were visible is that the gaze manipulation resulted in obscuring not only leakage gaze cues, but also cues that help coordination such as turn-exchange signals. While further experiments are required to disentangle the effects of leakage cues and coordination cues on performance, that the gaze manipulation affected the number of questions and not the time measure might suggest that the lack of coordination cues did not cause significant delays in the task and did not significantly affect participant performance.

5.3. Experiment II: Leakage Gaze Cues in Human-Robot Interaction and the Effects of Robot Design in the Perception of these Cues

The first experiment investigated whether people use information from others' gaze cues, including leakage gaze cues, to make inferences of mental states. The second experiment looked at whether people read and interpret correctly leakage gaze cues in two robots, Robovie and Geminoid, using the same guessing game experimental framework.

This experiment followed a two-by-two (two robots and "no gaze cue" vs. "leakage gaze cue" conditions), mixed-factorial design (robots as a between-participants manipulation and gaze as a within-participants manipulation) in which participants played the game with either one of the two robots in eight trials with an additional practice trial at the beginning of the experiment. In all of these trials, the robots played the role of the picker and participants played the role of the guesser (Figure 5.11). In half of these trials (excluding the practice trial), the robots produced leakage gaze cues before answering two of the first three questions that they received by glancing at their picks (as illustrated earlier in Figure 5.6). The robots' answers were delayed before questions in which they did not produce the gaze cue with the duration of the glance to keep the time it took them to answer questions consistent across trials and conditions. In summary, the two gaze conditions were as follows:



5.11. Robovie or Geminoid playing the picker and the participant playing the guesser in the guessing game.

No gaze cue condition – After the question, the robots waited (for the same amount of time that a glance took), looked up, established eye contact, and answered the participant's question.

Leakage gaze cue condition – After the question, the robots glanced at the object, looked up, established eye contact, and answered the participant's question.

Participants were randomly assigned to play the guessing game with one of the two robots. Each participant played four rounds of the game in each condition. The orders in which (1) the robot chose items and (2) the gaze manipulation appeared were counterbalanced. Except for the gaze manipulation, the robots' behaviors were identical across trials. To ensure that the two robots' gaze behaviors were designed to target the items on the table and be identical for the two robots, an accuracy test was conducted at the end of the experiment to assess the validity of the robots' gaze behaviors. This test evaluated how accurately participants could rate the gaze directions of the two robots and a human confederate as they glanced at randomly selected items on the table.

5.3.1. Hypotheses

Drawing from existing theory on nonverbal leakage (Argyle et al., 1971; Waxer, 1977; Feldman et al., 1978; Krauss et al., 1996), two main hypotheses were developed on how the gaze cue would affect participants' task performance and how the interpretation of the cue would differ between interactions with Robovie and those with Geminoid.

Hypothesis 1 – Participants will identify the item that the robots choose faster—using a smaller number of questions and spending less time—when the robots produce gaze cues than when they do not.[4]

Hypothesis 2 – The gaze cue will significantly affect task performance with Geminoid and not with Robovie. I predict that the leakage cue will be correctly interpreted with Geminoid but not with Robovie, as Geminoid's near-human features will facilitate the perception of the cue as a social signal and Robovie's stylized design will not do so.

5.3.2. Participant Sample

A total of 26 participants (17 males and 9 females) participated in the experiment. All subjects were native-Japanese-speaking university students recruited from the Osaka area. The ages of the subjects varied between 18 and 24 (*M*=20.4, *SD*=1.50). Participants represented a variety of university majors. Of the 26 participants, 11 studied engineering, 9 studied social sciences and humanities, 3 studied management sciences, 2 studied natural sciences, and 1 participant did not report university major. The computer use among participants was very high, averaging 6.50 (*SD*=0.65) on a scale from 1 to 7. Their familiarity with robots was relatively low (*M*=2.81, *SD*=1.55), as was their video gaming experience (*M*=3.00, *SD*=1.92) and online shopping experience (*M*=3.00, *SD*=1.52) on the same scale. One participant had a toy robot and 13 owned pets (8 dogs, 4 cats, and 1 ferret). A randomly selected 12 of these

---

[4] In the first experiment, the gaze manipulation did not significantly affect how much time it took participants to identify the item. I attribute this result to the between-participants design of the first experiment; the variance in the time measure casued by individual differences was greater than that caused by the gaze manipulation. For a within-participants design experiment such as the current one, I predicted that the time measure would also be significantly affected by the gaze manipulation.

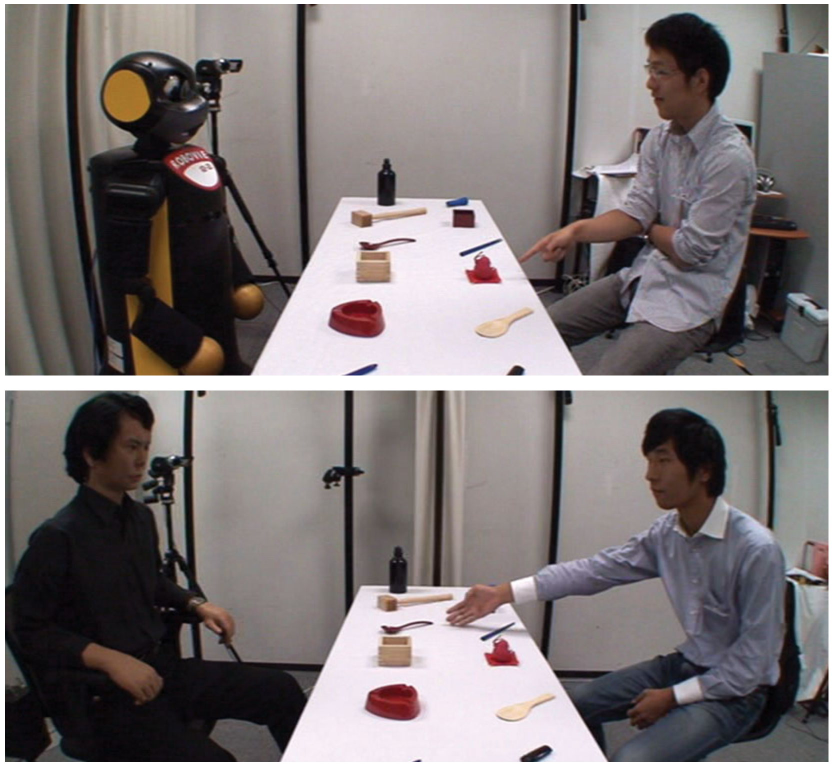


Figure 5.12. Robovie (top) and Geminoid (bottom) playing the picker and participants playing the guesser in the experiment.

participants (five males and seven females with an average age of 20.1, ranging between 18 and 22) were asked to take part in the accuracy test. Figure 5.12 shows participants playing the game with Robovie and Geminoid.

5.3.3. Experimental Procedure

Participants were first provided with a brief description of the purpose and procedure of the experiment. The experimenter told them that researchers at ATR have been designing robots that can play games with people and would like their help in testing their designs. The primary purpose of the experiment was concealed and participants were not given any information about the robots' behavior. After the introduction, participants reviewed and signed a consent form and filled in a pre-experiment questionnaire on their affective state. They were then provided with more detail on the experimental task. Participants were taken into the experiment room to play the game with either Robovie or Geminoid. After playing a practice round, they played eight rounds of the game. At the end of the game, the experimenter took them out of the experiment room and asked them to fill in a post-experiment questionnaire that measured their affective state, personality, perceptions of the robot in the experiment, overall experience with the experimental task, and demographic information. Finally, the experimenter interviewed all participants regarding their experience.

After all the subjective evaluations were obtained, a subset of the participants were asked to take part in a test that evaluated how accurately participants perceived the gaze directions of the two robots and a human confederate. To conduct this test, participants were taken back in the experiment room and asked to rate the gaze directions of Geminoid, Robovie, and a human confederate in a counterbalanced order in 12 trials with each gaze source.

The game task and the total experiment procedure took approximately 15 and 45 minutes, respectively. The experiment was conducted in a dedicated room with no outside distraction. The experimenter left participants in the room alone with the robots and observed the interaction remotely through live video feeds provided by two cameras. All subjects were paid ¥1,500 (approximately $14) for their participation including their travel expenses.

5.3.4. Measurement

The experimental design involved two manipulated independent variables: whether the robot produced the gaze cue (manipulated as within-participants) and whether participants played the game with Robovie or Geminoid (manipulated as between-participants). The dependent variables involved objective and subjective measurements.

Objective Measures – Participants' task performance was measured by capturing the time it took them and the number of questions they asked to identify the robots' picks. All sessions were videotaped to support the analysis of the objective measures.

Subjective Measures – Subjective measures evaluated participants' affective states using the PANAS scale (Watson et al., 1988), perceptions of the robots' physical, social, and intellectual characteristics using a scale developed for evaluating humanlike agents (Parise et al., 1998), attributions of mind and intentionality to the robots, perceptions of the task (e.g., how much they enjoyed and attended to the task), personalities using scales of intellectual competence, creativity, distrust, and empathy (Goldberg et al., 2006), and demographic information. Participants' affective states were measured before and after participants interacted with the robot and all other measurements were done after the experiment. Seven-point Likert scales were used in all questionnaire items. The post-experiment questionnaire included open-ended questions for a manipulation check; participants were asked to describe the cues that they observed in the robot's behaviors that helped them identify the picked item. The experimenter also conducted semi-structured interviews at the end of the experiment to gain a richer understanding of participants' experiences with and perceptions of the robots. Finally, in the accuracy test, participants were asked to mark the item toward which they perceived the robots or the human confederate glance on a questionnaire with a graphical representation of the items on the table. These markings were then compared to a list of the items that the robots and the human confederate were instructed to glance toward to obtain an accuracy rating for each gaze source.

5.3.5. Results

Objective measures were analyzed using a mixed-effects analysis of variance (ANOVA). Participant IDs were included in the model as a random effect and measured and manipulated independent variables (participant gender, pet ownership, gaze manipulation, and the robot with which participants interacted) and control variables (the target that the robot picked) were added as fixed effects. Subjective measures were analyzed using a fixed-effects analysis of variance. The accuracy ratings were analyzed using a random-effects, repeated-measures analysis of variance. Finally, the manipulation check was evaluated using a contingency analysis.

5.3.5.1. *Accuracy Check*

The accuracy test evaluated how accurately participants perceived the gaze directions of the two robots and a human confederate as they glanced at randomly selected items on the table. The analysis showed that participants rated Robovie's, Geminoid's, and the human confederate's glances at the items on the table with an average accuracy of 32.08% (SD=17.51%), 39.27% (SD=11.89%), and 37.50% (SD=15.28%) respectively with a baseline accuracy of 7.14% (for random guess). An analysis of whether participants rated the region of the item correctly, that is, rated either the exact item or one of its nearest neighbors, for the three sources showed accuracies of 85.00% (SD=10.68%), 78.98% (SD=16.54%), and 83.84% (SD=13.19%) Robovie's, Geminoid's, and the human confederate's gaze directions respectively. These results are also illustrated in the Figure 5.13. An analysis of variance comparing the ratings of the gaze directions of the robots and the human confederate showed that the overall model was not significant either for the exact item ($F[2,33]=0.63$, p=ns) or for the region of the item, $F(2,33)=0.92$, p=ns. Similarly, pairwise comparisons produced no significant differences among the accuracy ratings of the three gaze sources.

5.3.5.2. Objective Measures

Participant performance was assessed using two measures: the number of questions participants asked and the time it took them to identify the robots' picks. Data on these two measures were collected in 208 trials. Two of these trials were excluded due to operator error. The distributions of the two objective measures were transformed using the logarithm function.

The first hypothesis predicted that participants would perform significantly better in identifying the item when the robots produced the gaze cue than when they did not. Analysis of variance of the time measure provided full support and that of the number

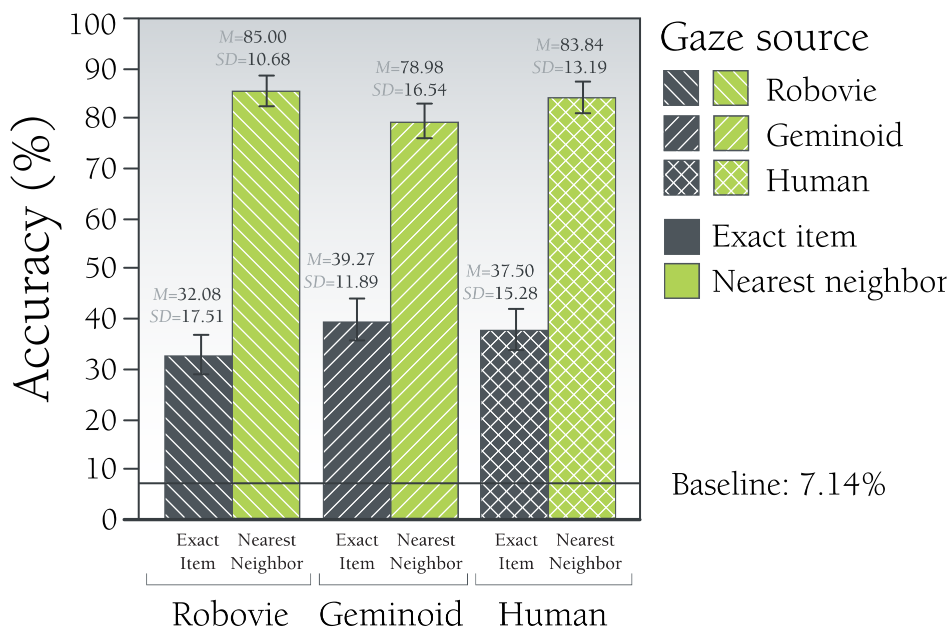


Figure 5.13. Accuracy of perceptions of Robovie's, Geminoid's, and a human confederate's gaze directions for the exact item and one of its nearest neighbors.

of questions measure provided partial support for this hypothesis. Participants took significantly less time ($F[1,167]=5.81$, $p=0.02$) and asked marginally fewer questions ($F[1,170]=2.81$, $p=0.10$) to identify the robots' picks when the robots produced the gaze cue than when they did not do so (Figure 5.14).

The second hypothesis predicted that the gaze cue would affect participant performance with Geminoid but not with Robovie. The analysis of variance did not find an interaction effect between robot and gaze manipulation over the time measure ($F[1,167]=0.17$, p=ns) or the number of questions measure, $F[1,168]=0.92$, p=ns. However, post-hoc pairwise contrast tests provided partial support for this hypothesis. Participants who played the game with Geminoid found the item significantly faster ($F[1,167]=3.93$, $p=0.05$) and with marginally fewer questions, $F(1,169)=3.42$, $p=0.07$. On the other hand, the gaze cue did not affect the performance of those who played the game with Robovie as measured by the time it took them to identify the item ($F[1,166]=2.05$, p=ns) nor did it affect the number of questions they needed to ask, $F(1,168)=0.27$, p=ns. The results from the pairwise contrast tests for the time measure are shown in Figure 5.14.

While the gaze cue affected the performance of participants who played the game with Geminoid but did not affect the performance of those who played with Robovie, a contingency analysis for the manipulation check (whether or not participants reported identifying the gaze cue and using this information to correctly guess the robots' picks) showed that significantly fewer participants reported identifying the gaze cue in Geminoid's behavior than in Robovie's behavior, $\chi^2(1, N=26)=7.54$, $p<0.01$. This result is shown in Figure 5.15. Furthermore, the analysis showed that those who reported identifying the gaze cue did not differ in performance from those who did not report

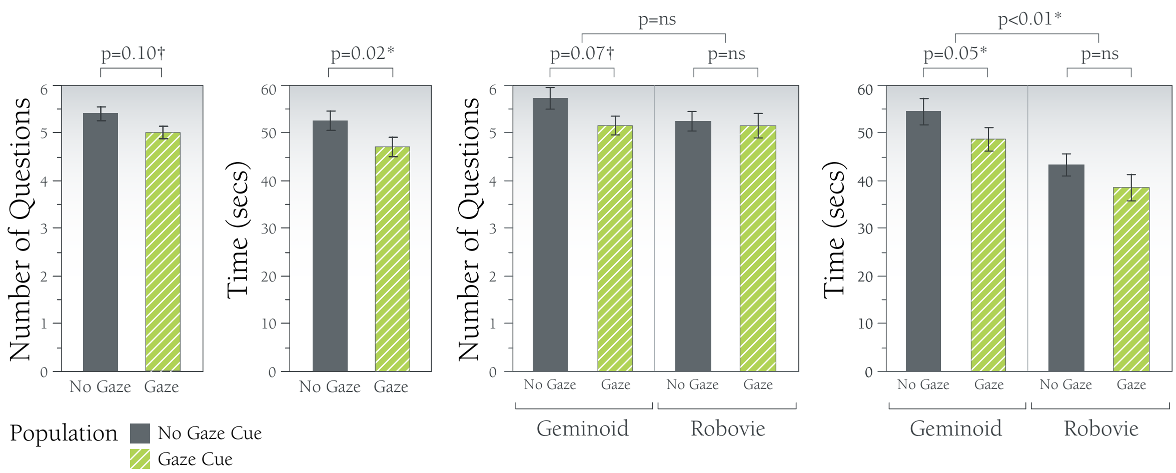


Figure 5.14. The number of questions participants asked to identify the item (left). The time it took them to identify the item (middle) and the same measure for each robot (right).

identifying the gaze cue in time (F[1,24]=0.77, p=ns) or number of questions, F(1,25)=0.23, p=ns. These findings are further supported by the qualitative data; several participants reported in the semi-structured interviews that they identified Robovie's gaze cues but did not attribute intentionality to the cue, which might explain why the gaze cue did not significantly affect their performance with Robovie. This explanation is further considered in the Discussion section.

The analysis also showed that, overall, participants identified the item significantly faster with Robovie than with Geminoid (F[1,23]=13.71, p<0.01), both when the robots produced the gaze cue (F[1,45]=8.21, p<0.01) and when they did not do so, F(1,46)=10.76, p<0.01. The number of questions that the participants asked was not affected by whether they interacted with Geminoid or Robovie, (F[1,24]=1.99, p=ns), either when the robots produced the gaze cue (F[1,102]=0.03, p=ns) or when they did not do so, F(1,99)=2.68, p=ns.

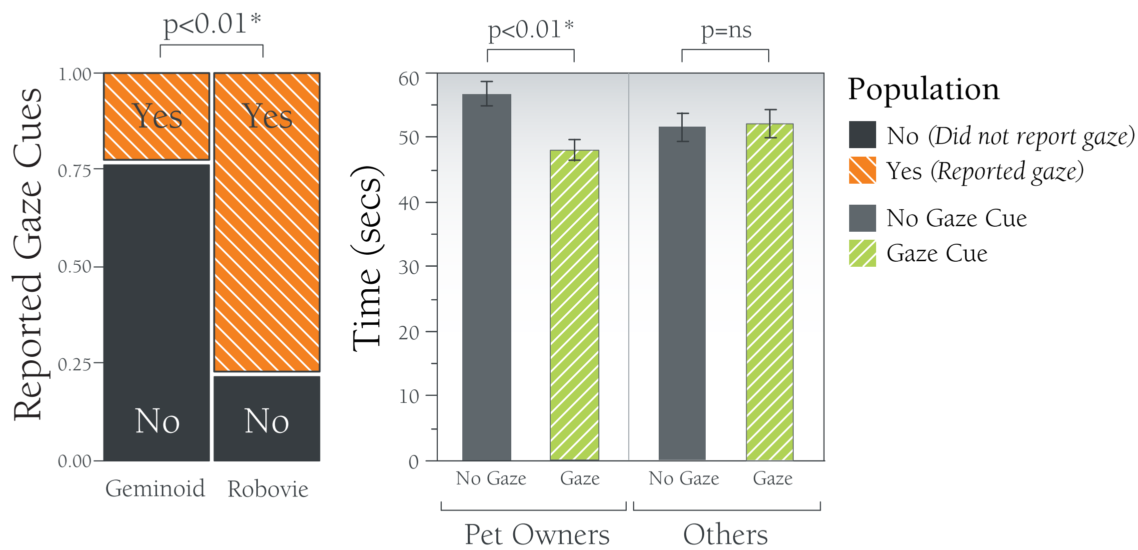


Figure 5.15. Whether participants reported identifying the gaze cues for each robot (left) and time it took pet owners and others to identify the item in gaze and no gaze conditions (right).

The analysis found no effect of gender on how the gaze cue affected participants' performance but found a significant interaction between pet ownership and the number of questions they asked to identify the robots' picks, $F(1,169)=4.52$, $p=0.03$. Those who owned pets identified the robots' picks using significantly fewer questions ($F[1,169]=7.30$, $p<0.01$) and in significantly less time ($F[1,166]=6.80$, $p=0.01$) when the robots produced the gaze cue. Those who did not own pets showed no differences in the number of questions that they asked ($F[1,169]=0.09$, p=ns) and the time it took them ($F[1,166]=0.64$, p=ns) to identify the robots' picks with the presence of the gaze cue. These pairwise comparisons are shown in Figure 5.15.

5.3.5.3. Subjective Measures

The analyses of the subjective measures first involved a factor analysis of 30 questionnaire items that were used to evaluate the social and intellectual characteristics of the robots. The factor analysis produced eight factors from which two reliable measures were created: a six-item scale of social desirability (Cronbach's α=0.84) and an eight-item scale of intelligence and attribution of mind (Cronbach's α=0.76).

An analysis of variance showed that participants rated Robovie as significantly more socially desirable ($F[1,24]=4.43$, $p=0.05$), more cooperative ($F[1,24]=7.06$, $p=0.01$), but less humanlike ($F[1,24]=10.54$, $p<0.01$) than they rated Geminoid (Figure 5.16). They also attributed marginally more intelligence and mental states to Robovie than to Geminoid, $F(1,24)=2.94$, $p=0.10$. This result is somewhat inconsistent with the qualitative data obtained through interviews; participants associated intelligence mainly with the fact that the robots could answer all of their questions in the game, and not with their physical design.

The analysis also found a significant interaction effect between participant gender and robot over social desirability, $F(1,22)=10.85$, $p<0.01$. Women rated Robovie as significantly more socially desirable than they rated Geminoid ($F[1,22]=17.03$, $p<0.01$), while no differences were found in men's ratings of the social desirability of the two robots, $F(1,22)=0.01$, p=ns. The analysis also produced a marginal interaction effect between pet ownership and robot over participants' ratings of the social desirability of the robots, $F(1,22)=3.72$, $p=0.07$. Those who did not own pets rated Robovie as more socially desirable than they rated Geminoid ($F(1,22)=8.76$, $p=0.01$) while pet owners' evaluations of the two robots did not differ $F(1,22)=0.05$, p=ns. These results are illustrated in Figure 5.16. Finally, a marginal interaction between attributions of intentionality and mental states and pet ownership was found, $F(1,22)=3.03$, $p=0.10$. Pet owners attributed more intentionality and mental states to

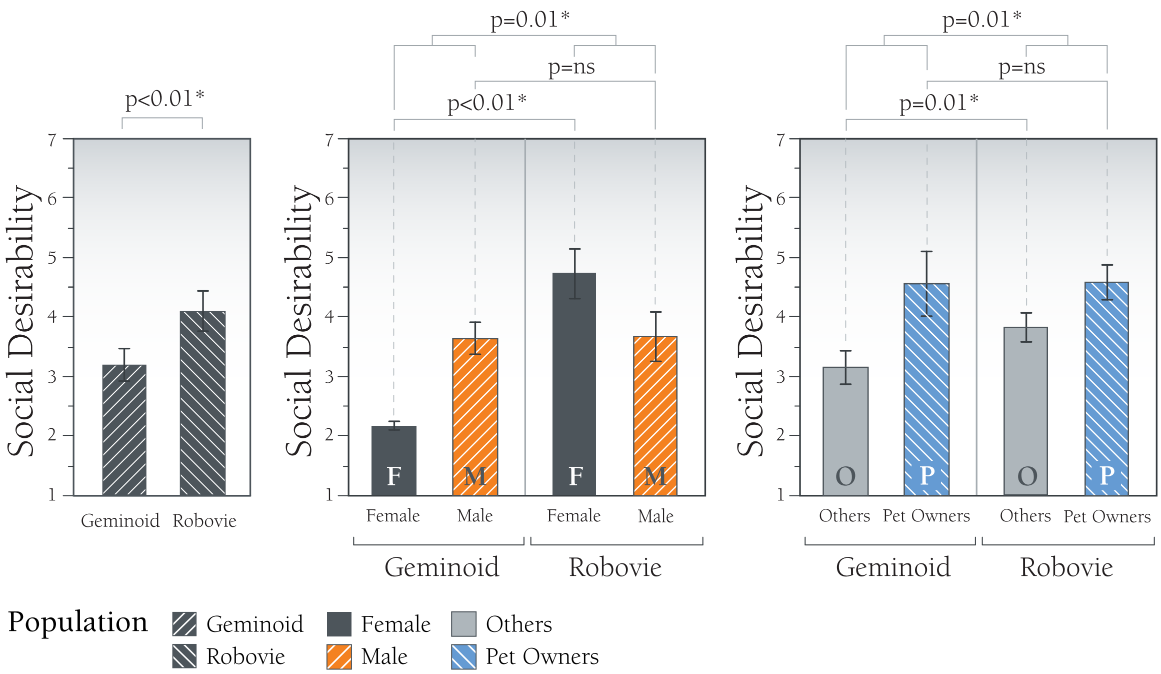


Figure 5.16. Ratings of the two robot's social desirability across all participants (left), for females and males (middle), and for pet owners and others (right).

Geminoid than they did to Robovie (F[1,22]=12.90, p<0.01), while non-pet owners did not show significant differences in their attributions, $F(1,22)$=1.28, p=ns.

Some of the factors in the factor analysis were loaded on single items. Therefore, single items were also analyzed using analyses of variance. No differences were observed in how much participants liked the robot ($F[1,24]$=0.02, p=ns) or how much they thought that the robot liked them, $F(1,24)$=0.85, p=ns. However, the analysis showed a significant interaction between robot and gender over how much participants liked the robot, $F(1,22)$=5.98, p=0.02. As a group, women liked Robovie more than they liked Geminoid ($F[1,22]$=4.74, p=0.05) and men did not show differences in how much they liked the two robots, $F[1,22]$=1.56, p=ns. Similarly, the analysis produced a significant interaction between robot and pet ownership over how much participants thought the robot liked them, $F(1,22)$=9.61, p=0.01. Those who did not own pets thought that Geminoid liked them significantly less than Robovie did ($F[1,22]=8.54$, $p=0.01$), while no differences were found in pet owners' ratings of how much they thought that the two robots liked them, $F(1,22)=2.14$, p=ns.

5.3.6. Discussion

The results supported the first hypothesis. Participants performed better on two performance measures when the robots leaked their mental states by means of gaze than when they did not. From this finding, it can be inferred that participants read the leakage cue, attributed a mental state to the robot, and used this information in their task. The second hypothesis was partially supported. Participants performed significantly better in the presence of the gaze cue when they played the game with Geminoid, but not when they played the game with Robovie. Also, participants were more likely to report identifying the gaze cue with Geminoid than with Robovie. Whether participants reported identifying the gaze cue did not affect their performance, supporting the argument that the perception of and responses to leakage cues can be automatic and subconscious.

The results also showed strong effects of pet ownership on all objective measures. Gaze cues affected only pet owners' performance in the game and not others, suggesting that people who own pets might become more sensitive to nonverbal behavior, as this is the main channel of communication between a pet and its owner. In support of this explanation, previous research found that dog owners learn to read the gaze cues of their dogs to understand their attentional states and needs (Miklósi et al., 2000). Developmental research has also shown that young males who live in households with pets show higher ability to decode nonverbal behavior (Guttman et al., 1983). Research on virtual agents has also shown that dog owners differed from others in how they evaluated agents with zoomorphic features (Parise et al., 1998).

The results also showed that, overall, participants performed better with Robovie than with Geminoid both when the robots produced leakage cues and when they did not. One explanation of this result is that interaction with Geminoid was cognitively and perceptually more demanding than interaction with Robovie was. Data from the semi-structured interviews provides some support for this explanation. Participants consistently reported being surprised by how humanlike Geminoid looked. They also reported feeling nervous, losing focus, and becoming distracted from their task. Two participants reported that they could not relate to the robot because “it looked older than them,” suggesting a cultural limitation and an alternative explanation for why participants performed more poorly with Geminoid than with Robovie. Participants also reported that their nervousness diminished over time, suggesting that allowing participants to interact with Geminoid in a non-intimidating task before they performed in the experiment might have alleviated some of the effects caused by the design of the robot.

In summary, both in the presence and absence of the gaze cue, participants performed better with Robovie than with Geminoid. I suspect that this effect was a product of Geminoid’s near-human appearance, which participants reported to be distracting. However, the effect of the gaze cue in improving participant performance was greater with Geminoid than with Robovie, even though fewer participants reported noticing the gaze cue in Geminoid than with Robovie. I argue that, though it was a distraction, Geminoid’s near-human appearance, in contrast with Robovie’s abstract design, led participants to more readily read the gaze cue (i.e., accurately determining gaze direction) and correctly interpret it (i.e., attribute intentionality and use this information to improve their performance in their task).

5.3.7. Limitations

The within-participants manipulation of the gaze cue limited the ability to measure the effect of the gaze cue on subjective evaluations of the robot. While this design allowed for accounting for some of the variability in participants' task performance that individual differences might cause, it fell short of providing insight into how leakage cues might affect subjective attributions of intentionality, purposefulness, and states of mind. This limitation is addressed in the third experiment of this study by introducing the gaze manipulation as a between-participants independent variable.

When the robots did not produce gaze cues, they delayed their answers by the same amount of time that it would have taken them to produce the leakage gaze cue. However, this artificial delay (an average of 1.12 sec) with no motion associated with it led to an unnatural pause during turn-exchanges. This limitation is addressed in the third experiment by recording precisely the time the robots took to produce the gaze cue and subtracting the recorded amount from the total task time.

5.4. Experiment III: Attributions of Intentionality to Leakage and Concealing Gaze Cues

The second experiment investigated whether people read and interpreted leakage gaze cues in two robots, Geminoid and Robovie, in a within-participants-design experiment. The third experiment described here studies this phenomenon in a between-participants-design experiment using the same experimental framework to understand how subjective evaluations of a robot might change in the presence of these cues and provide further evidence that leakage gaze cues lead to mental state attribution. Additionally, it investigates how people respond to concealing gaze cues in their attributions of mental states and evaluations of the robot.

The experiment followed a three-by-one, between-subjects design in which participants played the guessing game with Geminoid in one of the three conditions:

No gaze cue condition – The robot did not produce leakage or concealing gaze cues, providing only verbal responses.

Leakage gaze cue condition – The robot produced leakage gaze cues before answering two of the participants' first three questions, glancing at its pick.

Concealing gaze cue condition – The robot produced concealing gaze cues before answering two of the participants' first three questions, glancing at its pick and subsequently glancing at another item.

The robot's behavior was identical across trials, except for the gaze manipulation. To address the limitation caused by delaying the robot's responses in the second experiment, the robot's answers were not delayed in this experiment. Instead, the time it took the robot to produce the glances were recorded and subtracted from the total time it took participants to complete the task in the conditions where the glances were used. Participants played two practice rounds and ten recorded rounds of the game. They were randomly assigned to one of the three conditions. The order in which the robot chose the items on the table was counterbalanced.

5.4.1. Hypotheses

Drawing from results of the second experiment and from relevant literature, three hypotheses were developed:

Hypothesis 1 – Participants will identify the robot's pick using fewer questions when the robot produces leakage gaze cues, but their performance will not be affected when the robot produces concealing gaze cues.

Hypothesis 2 – Participants will attribute more intentionality to the robot when it produces leakage gaze cues than they do when it does not. Ratings of intentionality will be higher also in the presence of the concealing gaze cue than when the robot produces no cue.

Hypothesis 3 – Participants will rate the robot as less trustworthy and more deceptive when it produces concealing leakage cues than they do when the robot produces no gaze cues.

5.4.2. Participant Sample

Forty-eight college students, 28 males and 20 females, from the Osaka area participated in the experiment. Participants' ages ranged between 18 and 25 (M=20.77, SD=1.63). Of all the subjects, 32 studied social sciences and humanities, 10 studied engineering, 3 studied natural sciences, 2 studied management sciences, and 1 studied art. Participants rated their computer use as very high, averaging 6.54 (SD=0.80) on a scale from 1 to 7. Their ratings of their own familiarity with robots, video game experience, and online shopping experience were moderate, being on average 2.92 (*SD*=1.69), 3.38 (*SD*=2.13), and 2.96 (*SD*=1.88) respectively. Two participants owned toy robots and 12 participants owned pets.

5.4.3. Experimental Procedure

The experiment followed a procedure similar to that of the second experiment with two main differences. First, participants played two practice rounds of the game instead of one. The goal of this change was to alleviate some of the discomfort participants experienced in interacting with Geminoid by allowing them to gain more familiarity with Geminoid. Second, the experimenter entered the room at the end of the practice rounds and answered any questions that the participants had about the items on the table or their interaction with the robot. The goal of this change was to allow participants to ask questions about the items on the table after they play practice rounds to address any ambiguities that arose about the properties and functions of the items.

5.4.4. Measurement

The experiment had a single manipulated independent variable: whether the robot produced (1) no gaze cues, (2) leakage gaze cues, and (3) concealing gaze cues. The dependent variables were evaluated by objective and subjective measures.

Objective – As in the second experiment, two objective measures assessed participant performance: (1) time it took participants to identify the robot's pick, and (2) the number of questions they needed to ask to do so.

Subjective – In addition to the scales used in the second experiment, a post-experiment questionnaire assessed participants' attributions of mind and intentionality to the robot using a scale developed to evaluate people's judgments of the intentionality of others' actions (Malle & Knobe, 1997). All questionnaire items used seven-point Likert scales. As in the second experiment, a manipulation check was done using open-ended questions in the post-experiment questionnaire that explicitly asked participants to list the kinds of cues that they observed in the robots' behavior when identifying the robots' picks.

Qualitative – The experimenter interviewed participants to further investigate whether they recalled seeing the robot produce gaze cues and to gain a richer understanding of their perceptions of the robot.

5.4.5. Results

Objective and subjective measures were analyzed using a mixed-effects analysis of variance (ANOVA). Condition ID was nested under participant ID and included in the model as a random effect. The trial number and the ID number of the robot's pick were used as fixed effects in the analysis of the objective measures to control for effects of learning and difficulties participants might have had with identifying particular items. The manipulation check used counts of whether participants identified the gaze cues.

5.4.5.1. Objective Measures

As in the second experiment, participant performance was evaluated using two measures: (1) the number of questions they asked to identify the robot's pick, and (2) the time it took them to do so. The task performance data included 480 trials, 2 of which were removed due to operator error. The distributions of the performance measures were transformed using the logarithm function.

No gaze cue vs. leakage gaze cue – The first hypothesis predicted that participants, as they did in the second experiment, would read the leakage cue, interpret it as related to their task, and use this information to perform better in the game, which can be predicted in less time and by fewer numbers of questions. The analysis of the number of questions measure fully supported and the time measure partially supported this hypothesis; participants fro whom the robot produced leakage gaze cues asked significantly fewer questions ($F[1,31]=8.76$, $p<0.01$) and took marginally less time ($F[1,31]=3.93$, $p=0.06$) than those with whom it did not.

The analysis showed a significant interaction between gaze manipulation and participant gender over the number of questions, $F(1,29)=5.31$, $p=0.03$. Post-hoc analyses showed a similar trend in the time measure. Male participants asked significantly fewer questions ($F[1,29]=14.45$, $p<0.01$) and took significantly less time to identify the robot's pick ($F[1,29]=6.23$, $p=0.02$) when the robot produced leakage cues than when it did not. Female participants did not differ in the number of questions they asked ($F[1,29]=0.02$, p=ns) and the time it took them to identify the

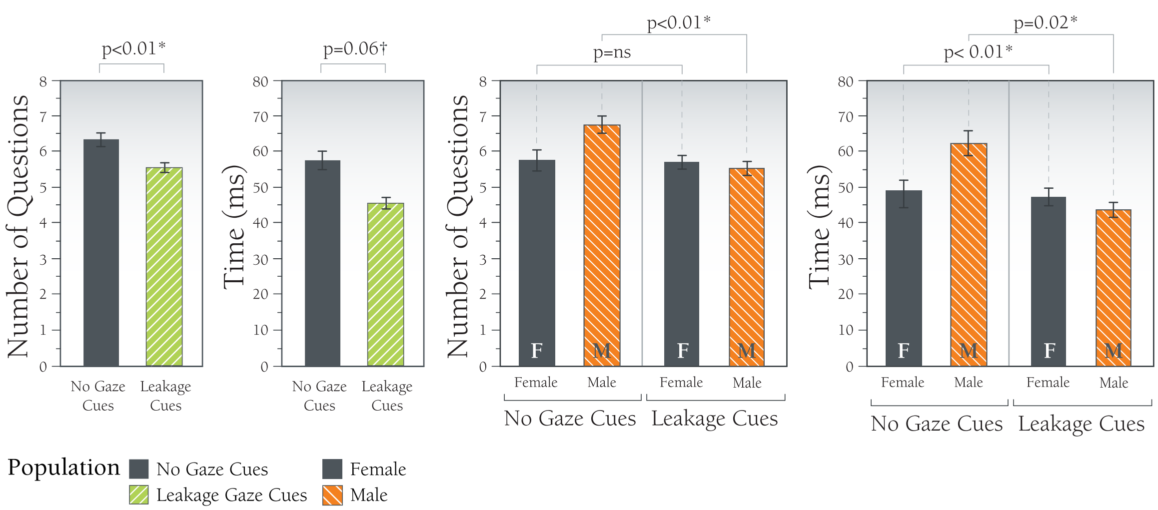


Figure 5.17. Number of questions and time measurements (left and middle-left) and same measures for females and males (middle-right and right) between no gaze cue and leakage gaze cue conditions.

robot's pick (F[1,29]=0, p=ns) when the robot produced leakage cues than when it did not. These results are illustrated in Figure 5.17.

No gaze cue vs. concealing gaze cue – The first hypothesis also predicted that, when the robot produced concealing gaze cues, it would indeed "conceal" the leaked information, and, therefore, participant performance would not show significant differences between no gaze cue and concealing gaze cue conditions. Results confirmed this hypothesis in both measures of performance. Participants did not differ in the number of questions they asked (F[1,27]=0.14, p=ns) and the time they took to identify the robot's pick (F[1,27]=0.17, p=ns) between when the robot produced concealing gaze cues and when the robot did not.

The analysis showed a significant interaction between the gaze manipulation and participant gender over the number of questions they asked, F(1,25)=7.20, p=0.01. Men asked marginally fewer questions (F[1,25]=3.88, p=0.06) when the robot

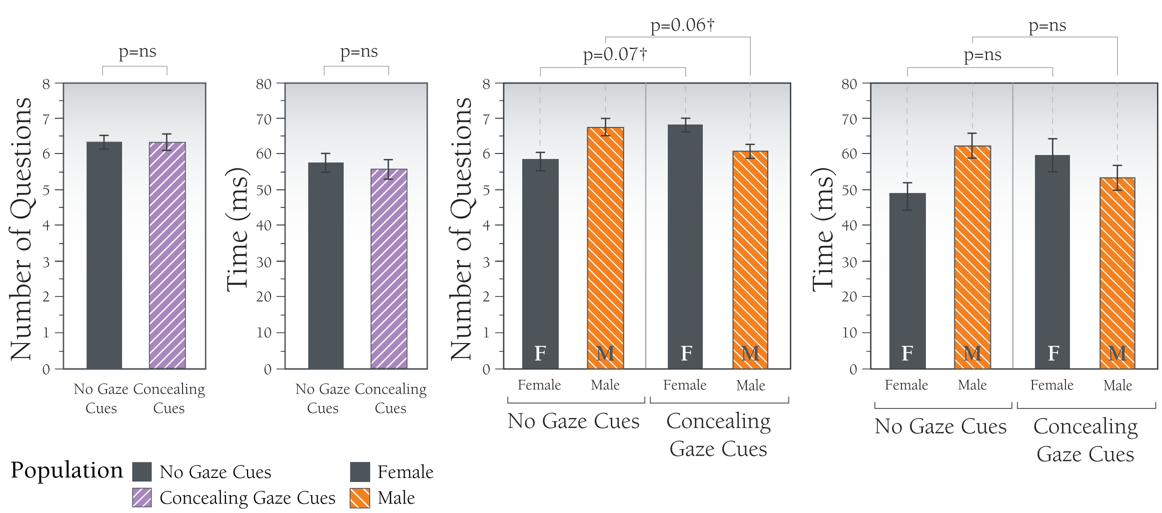


Figure 5.18. Number of questions and time measurements (left and middle-left) and same measures for females and males (middle-right and right) between no gaze cue and concealing gaze cue conditions.

produced concealing gaze cues than when it did not, while women asked marginally more questions when it produced concealing cues than when it did not, $F(1,25)=3.49$, $p=0.07$. Figure 5.18 illustrates results from the objective measures.

5.4.5.2. Subjective Measures

The analysis of the subjective measures included a factor analysis of the 41 questionnaire items that were used to evaluate social and intellectual characteristics of the robot. Eight factors were produced, from which four reliable measures were created: a seven-item scale of intentionality (Cronbach's $\alpha=0.84$), a six-item scale of rapport (Cronbach's $\alpha=0.81$), a four-item scale of sociability (Cronbach's $\alpha=0.82$), and a four-item scale of deceptiveness (Cronbach's $\alpha=0.76$).

No gaze cue vs. leakage gaze cue – The second hypothesis predicted that participants would attribute more intentionality when the robot produces leakage gaze cues than they would when it does not. The results did not confirm this hypothesis; participants' attributions of intentionality to the robot were not different when the robot produced leakage gaze cues than when it did not, $F(1,31)=0.08$, p=ns.

The analysis showed significant interactions between participant gender and gaze manipulation over several scales of subjective evaluations, particularly ratings of the robot's sociability ($F[1,29]=5.98$, $p=0.02$) and the game experience, $F(1,29)=6.83$, $p=0.01$. Post-hoc analyses showed that men rated the robot to be significantly less sociable when it produced leakage gaze cues than when it did not ($F[1,29]=9.66$, $p<0.01$) while women did not show differences in their evaluations across conditions, $F(1,29)=0.37$, p=ns. On the other hand, women rated their overall game experience to be significantly less positive when the robot produced leakage gaze cues

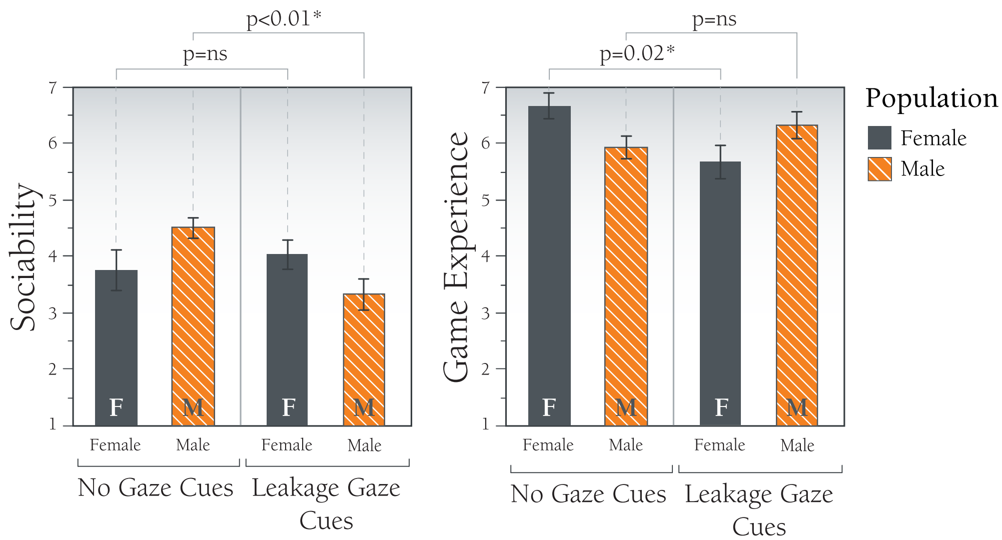


Figure 5.19. Subjective evaluations of the robot's sociability (left) and the overall game experience (right) for women and men in no gaze cue and leakage gaze cue conditions.

($F[1,29]=5.87$, $p=0.02$) while men showed no differences in their evaluations across conditions, $F(1,29)=1.37$, p=ns. These results are illustrated in Figure 5.19.

No gaze cue vs. concealing gaze cue – The second hypothesis also predicted that more intentionality would be attributed to the robot when it produced concealing gaze cues than when it did not. This prediction was not supported by the results. In fact, those with whom the robot produced concealing gaze cues rated the robot marginally less intentional than those for whom the robot produced no gaze cues, $F(1,27)=3.17$, $p=0.09$.

The third hypothesis predicted that participants would rate the robot as more deceptive when it produced concealing gaze cues than when it produced no gaze cues. The analysis found that ratings of the robot's deceptiveness did not differ between the no gaze cue and concealing gaze cue conditions, $F(1,45)=0.04$, p=ns. However, the cooperativeness scale provided partial support for this hypothesis; participants for

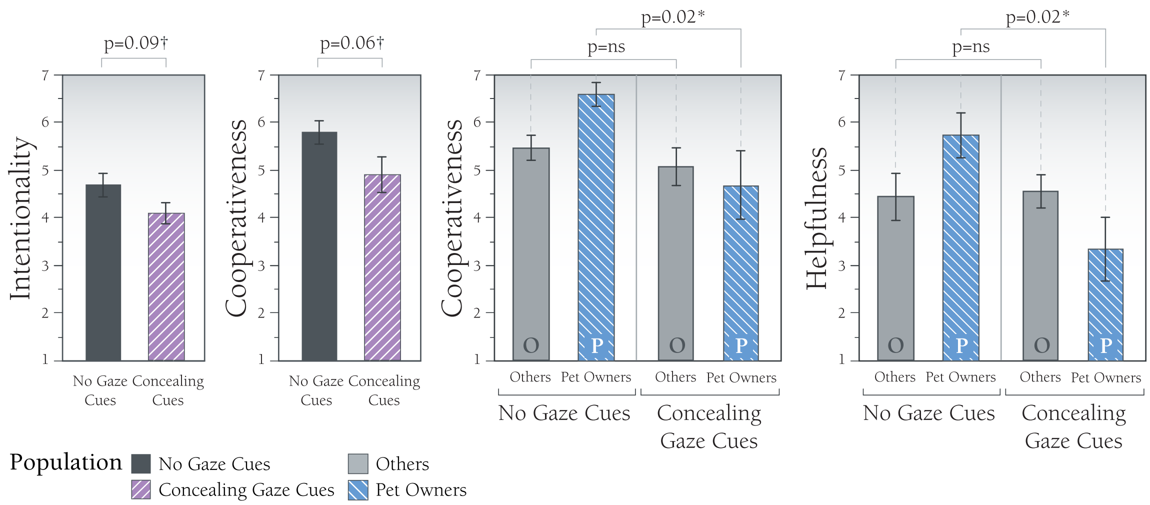


Figure 5.20. Subjective evaluations of the robot's intentionality (left), cooperativeness (middle-left), cooperativeness for pet owners and others (middle-right), and helpfulness for pet owners and others (right) in no gaze cue and concealing gaze cue conditions.

whom the robot produced concealing gaze cues rated the robot marginally less cooperative than those for whom the robot did not, $F(1,27)=3.82$, $p=0.06$. Post-hoc analyses showed that this effect was significant for pet owners ($F[1,25]=6.30$, $p=0.02$) and not for others, $F(1,25)=0.45$, p=ns. Similarly, the analysis found an interaction effect between pet ownership and gaze manipulation over how helpful the robot was, $F(1,25)=4.68$, $p=0.04$. Pet owners perceived the robot to be significantly less helpful when the robot produced concealing gaze cues than when it did not ($F[1,25]=6.74$, $p=0.02$), while others' evaluations of the robot's helpfulness did not change across conditions, $F(1,25)=0.02$, p=ns. Figure 5.20 illustrates these results.

The analysis of the subjective measures showed significant interactions between pet ownership and the gaze manipulation across several scales of subjective evaluation, particularly participants' evaluations of the robot's sociability ($F[1,25]=4.78$, $p=0.04$), the robot's intelligence ($F(1,25)=6.32$, p=0.02), their rapport with the robot

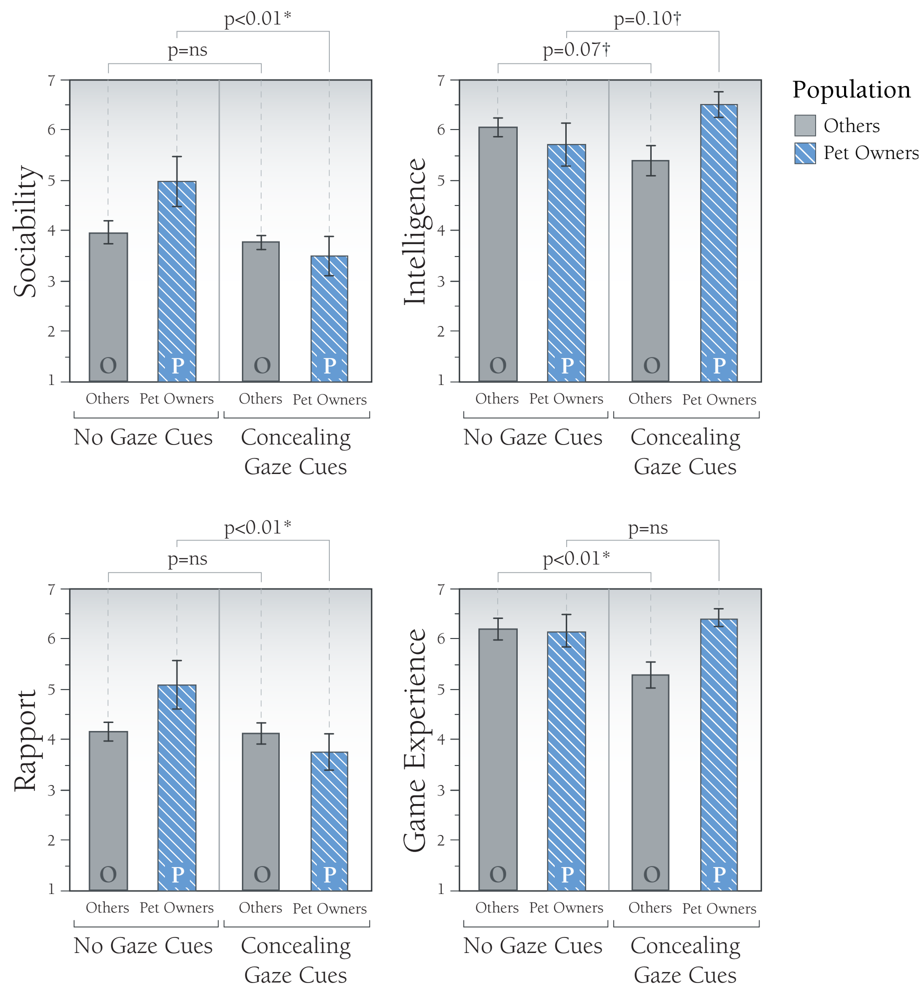


Figure 5.21. Subjective evaluations of the robot's sociability (top-left), intelligence (top-right), rapport with the robot (bottom-left), and game experience (bottom-right) for pet owners and others in no gaze cues and concealing gaze cues conditions.

($F[1,25]$=4.88, p=0.04), and the game experience, $F(1,25)$=5.42, p=0.03. Post-hoc analyses showed that pet owners found the robot significantly less sociable when it produced concealing gaze cues than when it did not ($F[1,25]$=9.68, p<0.01), while others did not differ in their evaluations, $F(1,25)$=0.24, p=ns. On the other hand, pet owners found the robot to be marginally more intelligent when it produced the concealing gaze cues than when it did not ($F[1,25]$=2.97, p=0.10), while others found the robot to be marginally less intelligent when the robot produced concealing gaze cues than when it did not, $F(1,25)=3.61$, $p=0.07$. Pet owners also reported significantly less rapport with the robot when it produced concealing gaze cues than when it did not ($F[1,25]=8.11$, $p<0.01$), while others did not differ in their evaluations, $F(1,25)=0.01$, p=ns. On the other hand, pet owners did not differ in their evaluations of their game experience ($F[1,25]=0.48$, p=ns), while others reported their experiences as significantly less positive when the robot produced concealing gaze cues then when it did not, $F(1,25)=8.65$, $p<0.01$. These results are illustrated in Figure 5.21.

5.4.5.3. Qualitative Observations

In the open-ended questions presented in the post-experiment questionnaire and the semi-structured interviews conducted at the end of the experiment, participants commented on the robot's behavioral characteristics, whether they identified the robot's gaze cues, and how they interpreted these cues. A number of participants reported mistaking Geminoid for a human confederate at first, feeling disturbed when they realized that it was a robot, mostly due to the robot's facial expressions (or lack of thereof), and not formerly having interacted with the robot prior to the experiment. Participants who identified the robot's gaze cues said that they tried to find a relationship between the robot's direction of gaze and its pick. While some found this information to be useful in finding the robot's pick, others thought that the two pieces of information were not related. In particular, those with whom the robot produced concealing gaze cues thought that the robot was "faking" or the robot was looking "randomly."

5.4.6. Discussion

The results supported the first hypothesis. Participants with whom Geminoid produced leakage gaze cues performed better in guessing the robot's pick than those with whom the robot did not, from which I infer that the participants read the leakage cue, attributed mental states to the robot, and used their attributions in their task. The results also supported the prediction that participants with whom the robot produced concealing gaze cues would not show improved performance, which suggests that the robot successfully "concealed" its pick by glancing at a randomly selected item on the table subsequently after producing a leakage gaze cue—a strategy that human participants frequently used.

Testing the second and third hypotheses provided further insights into how participants' subjective evaluations might be affected by robot's production of leakage and concealing gaze cues. The second hypothesis predicted that participants would attribute more intentionality to the robot when it produces leakage and concealing gaze cues than they do when it does not produce them. The results did not support this hypothesis. One explanation is that the intentionality scale failed to measure participants' attributions to the robot's production of the gaze cues. Another explanation is that participants interpreted "intentionality" as conscious, deliberate actions and both kinds of gaze cues were interpreted as unintentional acts. This explanation is further supported by the result that participants attributed marginally less intentionality to the robot when it produced concealing gaze cues than they did when it did not. Another support for this explanation is provided by participant's evaluations of the robot's fairness in the game. They rated the robot as significantly more "fair" when it produced the leakage gaze cues ($F[1,45]=4.02$, $p=0.05$) than when it did not. Similarly, they rated the robot as significantly more fair when it produced concealing gaze cues ($F[1,45]=5.27$, $p=0.03$) than when it did not. This result can be interpreted as participants attributing more fairness to the robot when it did not "intentionally withhold information" by not producing any gaze cues. However, further experimentation is needed to provide a more conclusive understanding of the relationship between gaze cues and attributions of intentionality.

The third hypothesis suggested that the concealing gaze cues would be associated with deceptiveness and, therefore, participants would rate the robot more deceptive when it produced concealing gaze cues than when it did not. The results provided partial support for this hypothesis; participants rated the robot as marginally less cooperative when it produced concealing gaze cues. Further analyses showed that only pet owners rated the robot as less cooperative when it produced concealing gaze cues. These individuals also rated the robot as less helpful when it produced concealing gaze cues, while others' ratings did not change across gaze conditions.

Gender and pet ownership had strong effects on participants' subjective evaluations of leakage and concealing gaze cues. Men found the robot to be less sociable when it produced leakage gaze cues, while women's evaluations of the robot did not change across gaze conditions. Pet owners rated the robot as more intelligent, but less sociable, and built less rapport with the robot when it produced concealing gaze cues than when it did not.

The first and the third experiments found effects of the gaze manipulation on the number of questions measure but not on the time measure. The second experiment found these effects on both measures. I attribute these differences to the different experimental designs of the three experiments. The gaze manipulation was introduced as a between-participants independent variable in the first and third experiments and as a within-participants manipulation in the second experiment. I argue that the length of time participants took to identify the item was greatly affected by individual differences, causing high variability in the time measurement. When this variability is controlled by a within-participants design such as in the second experiment, the effects of the gaze manipulation on the time measure could be identified. Therefore, I argue (and future work should consider the possibility) that number of questions might be a more robust measure of cognitive activity led by mental state attribution than time is.

5.4.7. Limitations

An important limitation in both the current and previous human-robot interaction experiments is that the design of the behavioral mechanism for the robots was limited to gaze cues that communicated mental states and a turn-taking mechanism (that the robot followed when answering participants' questions). Ideally the robot should have followed other gaze mechanisms such as the gaze patterns of an oratory during the greeting and leave-taking and gaze breaking before answering questions. However, I intended to keep the focus of the study narrow to answer a fundamental question: can we design gaze cues for a robot that would lead to attributions of mental states? Future work should examine how gaze cues that communicate mental states can be used as a communicative mechanism along with other behavioral mechanisms to construct more complex behavioral patterns.

In this experiment, the robot did not delay its answers in the no gaze cue condition to control for the time the robot took to produce gaze shifts in the other conditions. Instead, these times were recorded and subtracted from the total time. An alternative method for controlling this delay would be to have the robot produce a "gaze breaking" cue before answering the questions in the no gaze cue condition. However, this method was not technically feasible due to the mechanical limitations of the robot as the controller for Geminoid's eyes did not allow the robot's gaze toward targets to be higher than eye level—one of the directions that people look when they break gaze before answering questions.

An important limitation of all the experiments in this study is that they explored a particular kind of social cue in an extremely limited task context. Whether these results would generalize to other social cues and social situations is unknown. Because the context of the interaction plays an extremely important role in decoding these cues, future work should study leakage cues in a variety of social and task contexts and explore how these cues might be designed for robots specifically for these contexts.

5.5. Study Conclusions

Human communication involves a number of nonverbal cues that are produced unintentionally and communicate a wealth of information about the mental state of individuals. Leakage cues are a particular set of such cues that "leak" information about mental and emotional states through the nonverbal channel. This study explored whether people could read leakage cues, particularly leakage through gaze cues, in humanlike robots and make attributions of intentionality—that the robot has intentions or beliefs about the information that is leaked and how these cues might be designed through gaining a computational understanding of human behavior.

The first experiment looked at whether people used gaze cues to interpret others' mental states and found that participants performed better in a guessing game when they could see their partners' gaze cues. From this result I infer—with limitations that are discussed above—that they read their partners' leakage gaze cues, interpreted these cues as related to their task in the guessing game, and used this information to perform better in their task. The gender configuration of the dyads had an effect on their performance. All-female and all-male dyads showed the best and worst performance respectively. This ranking corresponds to the rankings reported in gaze literature on the total amount of mutual gaze in dyads, from which I speculate that increased total mutual gaze might lead to stronger perceptions of leakage gaze cues and attributions of mental states.

The second experiment investigated whether participants attributed mental states to robots and performed better in the guessing game when they produced leakage gaze cues and compared two robots with different levels of humanlikeness, Geminoid and Robovie. The results showed that participants performed better when the robots leaked information through cues as minimal as two short glances. I infer that they interpreted these cues to be related to their task and used this information to improve their performance in guessing the robots' picks. However, leakage gaze cues led to better performance when participants played the game with Geminoid but not when they played with Robovie, which might suggest that more humanlike physical features better support subtle cues. On the other hand, fewer participants reported identifying the leakage cue with Geminoid than with Robovie, suggesting a more automatic and subconscious response to the cues produced by Geminoid than those by Robovie. Furthermore, whether they reported identifying the gaze cue did not affect their performance, further supporting the argument that people automatically and subconsciously read and respond to leakage cues. Additionally, the leakage cue affected the performance of only pet owners and not others, which might suggest that those who own pets are more sensitive to nonverbal behavior.

The third experiment addressed some of the limitations of the second experiment and showed through a between-participants comparison that leakage gaze cues can communicate mental states of a robot and affect participant performance and subjective evaluations of the robot. Participants found the robot to be fairer when it leaked information. Some of these subjective evaluations were affected by participant gender. For instance, men found the robot to be less sociable when it produced leakage gaze cues, but the presence of these cues did not change women's ratings. The study also showed that the robot can successfully "conceal" the information that is being leaked by subsequently glancing so as to suggest incorrect information; participant performance was not affected when the robot produced concealing gaze cues. On the other hand, participants perceived the robot to be less cooperative when it produced concealing gaze cues. Pet ownership affected perceptions of the concealing gaze cue; pet owners found the robot to be more intelligent, but less sociable, and built less rapport with the robot. Table 5.1 summarizes the research questions, experimental designs, hypotheses, and results for all three experiments.

This study also has a number of research and design implications for human-robot interaction. Nonverbal leakages and, more broadly, seemingly unintentional behavior might provide designers with a rich design space for creating humanlike behavior. For instance, fidgeting might communicate nervousness more expressively than explicit facial or verbal expressions. However, it is important to note that the social context of the interaction will play a crucial role in how these cues are interpreted; the fidgeting might be interpreted as nervousness in one social context and hardware malfunction

Experiments	Results
Experiment I *– Do people use human leakage gaze cues to infer mental states?*	
Design – Two-by-one; no gaze vs. gaze as between participants; human-human interaction; guessing game task	
Hypothesis – Guessers would perform significantly better in guessing the pickers' picks when pickers' gaze cues are visible to them.	Supported
Experiment II *– Do people use robot leakage gaze cues to infer mental states? How does the design of the robot affect these inferences?*	
Design – Two-by-two; no gaze vs. gaze as within participants and Robovie vs. Geminoid as between participants; human-robot interaction; guessing game task	
Hypothesis I – Guessers would perform significantly better in guessing the robots' picks when the robots' produce leakage gaze cues.	Supported
Hypothesis II – The leakage gaze cue would lead to better performance with Geminoid and not with Robovie.	Supported
Experiment III *– How do leaking and concealing gaze cues affect inferences of mental states and attributions of intentionality?*	
Design – Three-by-two; no gaze, leakage gaze, vs. concealing gaze as between participants; human-robot interaction; guessing game task	
Hypothesis I – Guessers will perform better with leakage cue but not with concealing cues.	Supported
Hypothesis II – Participants will attribute more intentionality to the leakage cue and higher to the concealing cue.	Not Supported
Hypothesis III – Participants will evaluate the robot as less trustworthy when it produces the concealing cue.	Partially Supported

Table 5.1. Summary of research questions, experimental designs, hypotheses, and results for all three experiments.

in another. This study also informs research in shared attention and theory of mind in human-robot interaction. This study showed that even very short glances could lead to establishing shared attention, attribution of intentionality, and task performance effects. Furthermore, this study extends our understanding of how people interpret and respond to subtle human communicative cues when robots use them.

While this study provides evidence that gaze cues can communicate mental states of a robot and guidelines for how these cues might be designed, further work is required to generalize these results to a wider set of social contexts and to better understand how the design of the robots might shape people's judgments of nonverbal cues.

The next chapter discusses the limitations of the work presented in this dissertation, and draws on these limitations to provide a roadmap for future work.

6. General Discussion

In this dissertation, I attempted to address a highly complex and unconventional design problem—designing social behavior for humanlike robots—with the specific goal of achieving social and cognitive benefits. To address this design problem, I chose to take the approach of first gaining a deeper understanding of human social behavior and using this understanding to create the appropriate social behavior for humanlike robots. However, this choice raised the following question: Is this the best approach to designing social behavior? Section 6.1 attempts to answer this question.

To gain a better understanding of human social behavior from a design perspective and use this understanding to design social behavior for robots, I used knowledge and methods from a variety of research areas and made a number of design decisions on what knowledge and methods to use and how to use these resources. These decisions raise several questions regarding methodological validity. For instance, are the behavioral models that I created the best representations for the modeled behavior? Section 6.2 discusses these and other questions of methodological validity.

The experimental evaluation of the designed gaze behaviors showed a number of significant human social and cognitive outcomes led by manipulations in these gaze behaviors. However, these results were obtained in specific social and task settings, with specific populations, and using specific research platforms. Therefore, questions remain regarding the generalizability of these experimental results. Do the results presented here extend into other user populations, tasks and interaction scenarios, robotic platforms, agents in other modalities (e.g., virtual agents), or other nonverbal cues? How could more generalizability be achieved? These questions are addressed in Section 6.3.

Finally, in following this design process to create social behaviors for robots, I faced a number of technical and methodological challenges that remain significant bottlenecks in advancing the design of social behavior for robots. Section 6.3 discusses these central challenges and provides a vision for how future work might address them.

6.1. Design Approach

In my attempt to address the problem of how to design social behavior for robots, I chose to first understand *human* social behavior as a resource for designing humanlike robot behavior. This choice was inspired by a systems design perspective; social interaction, as with any other complex system, is made up of interrelated components and mechanisms that interact with each other, and, therefore, designing artificial elements to work with this system needs to be grounded in a deeper understanding of these components and mechanisms and the relationships among them. Whether the systems design perspective is the best approach to designing social behaviors for robots is still an open question. However, this dissertation showed that modeling human behavior from a design perspective reveals a number of design variables, mechanisms, and patterns that were previously unknown and unavailable for designers to create social behaviors that human communication system would appropriately reciprocate. While comparisons of the effectiveness of different design perspectives would provide a more conclusive answer, I argue that the complex nature of social interaction requires the design of artificial social stimuli to be grounded in a deeper understanding of the variables and mechanisms in this design space.

However, I also acknowledge that behavioral modeling is not the only way of capturing the design variables and mechanisms in social behavior. Other approaches that primarily rely on designer's intuition and guidelines developed through an iterative process—such as the "12 basic principles of animation" that Disney animators developed to create the "illusion of life" (Thomas & Johnston, 1981)—might be as effective. These approaches can be compared to the design approach presented here in future work. Furthermore, different design approaches might be more appropriate for different framings of the design problem. For instance, while the approach taken in this research might create behaviors that fit a robot with highly humanlike appearance, an animation artist's approach might create behaviors that are more appropriate and effective for a robot with an abstract design.

6.2. Methodological Validity

To make sense of the complexity in human social behavior and use it as a resource for creating robot social behavior, I developed and followed a design process that adapted methods and knowledge from a number of scientific research areas. In this process, I also made a number of decisions and judgments, such as choosing a particular method of analysis over another or focusing on certain design variables while omitting others, based mainly on my intuition and experience. While I sought ways to formalize some of these design decisions by grounding them in human communication theory or empirical results, I could not formalize and validate all of them given the complexity of the design problem and the limited time and resources of this dissertation research. The validity of these design decisions could be improved through seeking external validation at significant stages of the design process. However, because this process might involve hundreds or thousands of design decisions when working in a complex design space, the designer's intuition and experience will have to inform what decisions and analyses should be validated. Below, I discuss the limitations imposed by some of the decisions I made in conducting this research, and how future work could address these limitations.

6.2.1. Assumptions in Behavioral Modeling

Social behavior is an infinitely complex space for design variables and the relationships between and among them. In modeling gaze mechanisms, I made design decisions to focus on variables that I found to be most salient and important while omitting others. For instance, previous research has shown that the total amount that conversational partners look at each other decreases over the course of the conversation (Abele, 1986). The design of the conversational gaze mechanisms in the second study did not consider this variable as a part of the design as a means to simplify the design space. However, whether including this variable in the design might have changed the measured social and cognitive outcomes remains unknown.

In this complex design space, I also made a number of design decisions about the level of detail in modeling gaze behavior. The first study augmented an existing model of the relationship between gaze behavior and the sentence-level information structure of the spoken discourse. In the context of the second study, the literature suggested that sentence-level structures might not be the appropriate level of granularity for analysis and that phrase-level or topic-level structures might be used. I made the design choice of using topic-level structures as the basis for the gaze model. Ideally, the data could be modeled for both levels of information structure and tested for how well they could predict gaze behavior to choose the right level of granularity.

In creating the oratorial and conversational gaze models presented in the first and second studies, I did not take into account the majority of the addressees' behaviors (except modeling turn-exchanges and greeting and leave-taking rituals). While this is a common approach taken in research on gaze behavior, it falsely assumes that participants' behaviors are driven by internal states and are independent of the actions of their conversational partners. Duncan et al. (1984) argue that ignoring the contingency between the actions of conversational partners, what is called the "partner effect" (Kenny & Malloy, 1988), might cause substantial error in understanding interactional processes. Duncan et al. call this situation "pseudounilateriality," "the false assumption that the variable [e.g., how much a speaker looks at an addressee] is necessarily unilaterally determined by the actions of the participants." While the robots did not have the technical capability to capture and respond to their partners' actions in real time and account for them in generating their own behaviors, this assumption poses important limitations on the gaze models created using this process.

Future work can address these limitations by seeking ways to further formalize the design process and more rigorously study the design space. Such an approach would significantly improve the validity of design decisions. For instance, testing phrase-level and topic-level structures for the extent to which gaze shifts can be predicted would have provided empirical evidence for using topic-level structures in discourse. Furthermore, a more rigorous consideration of the variables in the design space would avoid biases such as the pseudounilateriality assumption. While it is important to note that a more thorough analysis of data would require extended resources, automating parts of the modeling process might facilitate analyzing a larger number of design variables. How techniques in speech and vision processing and data mining might be used to automate the modeling process is discussed in the next section.

6.2.2. Singling Gaze Out

One of the most important limitations of all three studies is that gaze was singled out from among the full set of nonverbal cues that compose visible behavior. In the first study, arm and body postures were used to enrich ASIMO's expressiveness as a storyteller, but, because adding these gestures might have confounded the results of the study, they were eliminated in the second and third studies. However, in human communication, a number of nonverbal mechanisms such as facial expressions, posture, and arm, head, and bodily gestures co-construct visible behavior along with spoken discourse. Furthermore, when highly humanlike research platforms such as Geminoid are used, the more subtle behaviors such as breathing, blinking of the eyes, fidgeting, and stretching might be required to create the impression of lifelikeness. The third study showed that participants performed worse with Geminoid than they did with Robovie, perhaps because these behaviors were not designed into the otherwise very humanlike robot. While the main focus of this dissertation was on designing behavioral mechanisms that can deliver social and cognitive outcomes, the results presented here suggest that a mismatch in lifelikeness between a robot's appearance and behavior might weaken these outcomes.

Future work should look at how different nonverbal behaviors could be combined to create visible behavior. For instance, body orientation plays an important role in communicating one's direction of attention and needs to be considered along with the head and eye movements that make up gaze behavior. Arm and hand gestures also play an important role in conversations, supporting the spoken discourse and communicating information that cannot be efficiently conveyed through speech (Chawla & Krauss, 1994; Cassell et al., 2007; Becvar et al., 2008). Future work should investigate how these cues might be integrated into designed behaviors to support the robot's gaze cues and speech. Human communication literature suggests a strong interaction between gaze and interpersonal distance (Argyle & Dean, 1965), and behavioral models should also consider the proxemic context of the social situation in generating gaze behaviors.

In the third study, participants reported discomfort with the lack of expressiveness in Geminoid's face. I posit that robots with the apparent physical ability to produce facial expressions will raise expectations of appropriate behavioral expressiveness and will need to produce the appropriate behaviors to meet these expectations. A consideration of whether these expectations are met would also improve the validity of the social and cognitive outcome measures.

6.2.3. The Use of Wizard-of-Oz Techniques

An important limitation of this research is the controlled interaction people had with robots. While designed gaze behaviors were implemented algorithmically and gaze was produced automatically and adaptively to robots' speech, other aspects of the interaction relied on the use of Wizard-of-Oz techniques. For instance, in all three studies, robots did not sense participants' locations or verbal responses. Instead, participants were seated at designated locations, and the robots were programmed to look at these locations. However, this technique did not account for the variability in participants' heights, which might have weakened the feeling of being looked at for participants who were much shorter or much taller than the average height that was considered in determining the robots' gaze target.

Robust vision and natural language processing techniques would help future work address these issues and allow the construction of a truly interactive experience for the participants. Furthermore, using specialized processing systems, such as a speech recognizer that is trained to recognize the spoken language used in any particular application domain or task of the study, might provide a near-term solution to avoiding the use of operators and designing more interactive experiences.

6.3. Generalizability

As with all experimental research in human-computer interaction, the results presented in this dissertation might not generalize beyond the cultural contexts and user populations of the empirical studies. The experimental scenarios and the robotic platforms used might also impose restrictions on the generalizability of these results. Furthermore, whether results that are obtained in a controlled experimental setting would extend into real-world contexts is not known with certainty. These factors are considered below.

6.3.1. Gender, Culture, and Language

Research on gaze in human communication suggests that gender has a significant effect on both the production and perception of gaze cues (Argyle & Ingham, 1972; Abele, 1986; Bente et al., 1998; Bayliss et al., 2005). The studies presented in this dissertation restricted the size and composition of studied populations, which in return limited the generalizability of the results. In the first study, the design of ASIMO's gaze behavior was based on a female storyteller in an all-female triad.

Whether the results can be replicated with a design based on a male speaker or with a female speaker in a triad with a different gender configuration is unknown. In the second study, both the design and evaluation of Robovie's gaze behaviors were based on male participants. The results of this study might only be generalized to male populations. Furthermore, the modeling of the leakage gaze cue in the third study only used male dyads. Whether the results could be replicated with a model obtained from all-female or female-male dyads is unknown.

Gaze behavior is also sensitive to cultural context and language. For instance, Ingham (1972, as described in Argyle and Cook, 1976) found significant differences in how much, how long, and how often Swedish and British participants looked at their partners during conversation. Designed behaviors and the social and cognitive outcomes that they lead to are limited to the cultural context and language of each particular study. In the fist study, ASIMO's gaze behavior was designed based on data collected from an English-speaking Icelandic storyteller and two English-speaking American addressees. Whether the gaze behaviors of a second-language speaker are different from those of a native speaker is unknown. Also, native English-speaking American participants were hired to evaluate the gaze behavior. Further experiments are required to understand whether the results from the experiment can generalize to other populations. Studies II and III involved all native-Japanese speakers for both the design and evaluation of the gaze behavior. Whether results from these studies can apply to non-Japanese populations needs further investigation.

Studies that compare results presented in this dissertation across cultures, languages, and user attributes (e.g., gender, age, personality, social status, and occupation) would significantly improve their generalizability. Future work should look at how designed behaviors could be extended to robots that work in different cultural contexts, use different languages, and interact with people with different demographic and personality attributes.

6.3.2. Experimental Tasks and Research Platforms

The tasks devised for the empirical studies also place some limitations on the generalizability of their results. For instance, the topic of conversation is found to affect how much people look at each other (Abele, 1986). The first study used storytelling as the context of the study. In the second study, Robovie provided travel information. In the third study, participants played a guessing game with Robovie and Geminoid. Whether the results from these studies would generalize to different tasks and conversation topics is unknown.

Another important limitation of this research is imposed by the physical and mechanical designs of the research platforms used in the studies. While these studies have shown with three different robots that humanlike gaze behavior can lead to social and cognitive outcomes that are predicted by human communication theory, whether these results would generalize to interactions with other robots is unknown. For instance, research on gaze has shown that the characteristics of a gazing confederate (e.g., gender) can significantly affect people's responses to the confederate (Patterson et al., 2002). Therefore, people's perceptions of the robot's characteristics might affect their responses to and perceptions of the robot. Powers and Kiesler (2006) showed that a robot's physical characteristics such as whether it had a male or female voice, the fundamental frequency of its voice, and the length of its chin predicted participants' rating of how knowledgeable and sociable they found the robot and whether they would follow health advice from the robot. Therefore, future work needs to test the generalizability of these results to interactions with other robots in order to gain a better understanding through systematic studies of how different characteristics of the robot might shape people's perceptions of and responses to the robot.

Further experiments are needed to understand whether these results generalize to other experimental scenarios and research platforms. Future studies that compare the social and cognitive impact of interaction in different social situations and with different robots would help to provide an understanding of how tasks, experimental scenarios, and the physical design of the robots affect human-robot interaction. Furthermore, looking into how much these findings might extend into other modalities (e.g., interactions through video and with on-screen agents) and levels of agency (e.g., interactions with autonomous agents and avatars) may pose a fruitful area for investigation.

6.3.3. The Controlled Laboratory Setting

The research approach presented here used controlled laboratory experiments to understand the social and cognitive outcomes of the designed gaze behaviors. This type of setting imposes important limitations on the generalizability of the results. Whether these outcomes could be obtained in less controlled environments is unknown. For instance, it is now known whether ASIMO's increased gaze will lead to greater information recall in a real-world classroom over longer periods of interaction.

To generalize the results beyond controlled laboratory settings, future work needs to also situate designed behaviors in real-world scenarios and contexts. For instance, testing whether a robot could use gaze cues to shape the conversational roles of its partners in a public environment such as a shopping mall with individuals who are not paid to interact with the robot can provide important insights into the generalizability of these results.

6.4. Technical Challenges

The research presented here also suffers from methodological and technical bottlenecks, particularly in modeling human behavior and creating real-time interactivity for robots. Applying techniques from areas such as speech and vision processing, data mining, machine learning, and databases to these problems might significantly help future work on designing social behavior overcome some of the limitations discussed earlier. Some of these methodological and technical obstacles and directions for future research are discussed below.

6.4.1. Modeling Social Behavior

The techniques and methods used in modeling human behavior also pose some limitations. For instance, in coding video data, all three studies used human coders, which puts limitations on the amount of data coded, the coding categories, and how biases and error can be introduced during the modeling process. Future work should look into automating this process using computer vision techniques and use estimations for missing data and error correction using semi-supervised machine learning techniques. Additionally, I used simple hierarchical probabilistic state machines to computationally represent these gaze models. While these representations might be sufficient to model the amount of human behavior data collected for the mechanisms considered in this work, future work should look into finding better ways to represent large amounts of data and to generate multiple sequential streams of events using techniques such as Hierarchical Hidden Markov Models or Conditional Random Fields.

Future work on analyzing human behavior and building computational representations will greatly benefit from exploring how research in computer vision, machine learning, data mining, and databases can automate analyses of human behavior data and find better computational representations. Computer vision techniques will be useful in automatically coding video data for specific behaviors. Data mining research can significantly improve the process of identifying structure in unstructured data, particularly in finding behavioral patterns in parallel streams of speech and nonverbal behavior data. Building more sophisticated computational representations for behavioral models will require processing and learning from large amounts of data. Database research can contribute to storing large amounts of data and testing hypotheses about the behavioral models by providing query interfaces. Finally, machine learning techniques will facilitate the process of finding patterns of co-occurrence in parallel, interdependent streams of behaviors, represent these patterns in temporal probabilistic frameworks, and provide real-time behavior generation. These techniques can also provide estimations for missing data and errors that occur in vision processing and data mining. Using these technologies will also facilitate studying complex interaction processes, for instance, taking into account all participants of a conversation in understanding the behaviors of a speaker, thus avoiding errors that might be caused by pseudounilaterality.

6.4.2. Real-time Interactivity

Due to the state of speech recognition and vision processing systems, today's robots offer very limited interactivity in generating behavior and constructing conversation. In the future advances in speech recognition and vision processing will allow researchers to create more interactive conversational mechanisms and applications.

Presently, however, until there are reasonable advances in these areas, limiting the conversational context (requiring that the robot recognizes words from a limited vocabulary set) and instrumenting the environment or users with sensors (to substitute or support the vision processing system) will minimize the error rates in speech and activity recognition and help the development of more interactive behavioral models and applications.

Building real-time interactivity into humanlike robots will require combining speech and nonverbal behavior recognition and generation and cognitive representations of the world that adapt to new input from users and the environment. Speech and nonverbal behavior recognition would significantly benefit from advances in natural language and vision processing. Similarly, speech and nonverbal behavior generation would benefit from building more sophisticated models of social behavior that use input from the environment, particularly from recognized speech and nonverbal behaviors of a partner. Furthermore, situationally aware representations of the real world need to process the input from recognition systems, make sense of the recognized input, and generate the appropriate response to it. Finally, these cognitive representations and behavioral models might be updated over time using unsupervised learning techniques. For instance, gaze research has shown great individual differences in how much people look at interaction partners[5]. A robot might need to adapt to these individual differences in recognizing and generating speech and nonverbal behavior.

[5] Nielsen (1962) reports the total time spent looking to range from 8% to 73% across individuals.

6.5. Summary

The work presented in this dissertation is a first step towards establishing a theoretical and methodological basis for designing behavioral mechanisms for humanlike robots and understanding their social and cognitive impact. However, a number of questions remain unanswered regarding the design approach, the validity of the methodology, the generalizability of the results, and the general methodological and technical obstacles in advancing the design of social behavior for humanlike robots. Comparisons of the effectiveness of different design approaches might cast light on whether using an understanding of human social behavior as the main resource for designing social behavior for robots is the best approach to addressing this design problem. The validity of the research approach could be improved by further formalizing the design process, seeking external validity at significant stages of the design process, and more rigorously studying the design space for possible interactions between design variables. Questions about the generalizability of the results could be answered through comparative studies that investigate how these results might extend into other user populations, platforms, modalities, and experimental scenarios. Field deployments and experiments could also provide a better understanding of how the results presented here might generalize to real-world contexts and to long-term social interactions. The generalizability of the research approach and process could be better understood by exploring whether they could be used to design other aspects of social behavior, particularly proxemic behavior, posture, and gestures. Finally, new tools and methods can facilitate modeling human behavior, finding better computational representations for communicative processes, and building systems that can learn from and respond to real-time input from the environment and interaction partners. My hope is that future work will explore these areas and build on this work.

7. Conclusions

The prevalent vision for humanlike robots is that by drawing on human physical, cognitive, and social capabilities they will one day provide us with significant social and cognitive benefits. Despite encouraging developments in robotics and increasing public interest, whether these systems *can* deliver these promised benefits has been greatly understudied. Furthermore, attempts to develop and systematically evaluate an interdisciplinary approach on *how* these systems can be designed so that they deliver these benefits has been extremely rare.

The goal of this dissertation is to develop an approach to designing social capabilities for humanlike robots, which draws on a theoretically and empirically grounded understanding of human social processes, and to demonstrate how these capabilities could deliver social and cognitive benefits through a series of empirical studies. Towards these larger goals, this work has made a set of methodological, theoretical, and practical contributions. The methodological contributions include an interdisciplinary, integrated process for designing, building, and evaluating social behavior for humanlike robots. These contributions are listed in Section 7.1. The theoretical contributions advance our understanding of human communicative mechanisms from a computational point of view and of people's responses to theoretically based manipulations in these mechanisms when they are enacted by humanlike robots. Section 7.2 summarizes these contributions. The practical contributions include the computational models of social behavior created for the empirical studies, which are described in Section 7.3. The last section in this chapter provides my closing remarks.

7.1. Methodological Contributions

This dissertation presents a unique process for studying and designing human communicative mechanisms and a demonstration of an interdisciplinary research approach that combines techniques and methods from communication research, discourse analysis, and computational linguistics to extract design variables from and create computational representations of human social behavior. This work also created a number of experimental paradigms in which these behavioral models were manipulated and evaluate through objective, subjective, and behavioral measures the social and cognitive outcomes of these manipulations. Table 7.1 lists these contributions.

Context	Contributions
All Studies	A theoretically and empirically grounded, interdisciplinary process for designing, building and evaluating communicative mechanisms for humanlike robots.
Study I	An experimental framework for studying how speaker attention could be manipulated through changes in gaze behavior and measuring the effects of different levels of attention on information recall and subjective evaluations of the speaker.
Study II	An experimental framework for studying how speakers could signal conversational roles through gaze cues and measuring whether people conform to these roles and the effects of conforming to these roles on information recall, task attentiveness, liking, and feelings of groupness.
Study III	An experimental framework for studying leakage gaze cues in human communication and human-robot interaction and measuring how the presence and absence of these cues might affect attributions of mental states using task performance measures.

Table 7.1. Methodological contributions of the dissertation.

7.2. Theoretical Contributions

The theoretical contributions of this work consist of two sets of new knowledge; a deeper understanding of human gaze mechanisms as applied to robots and their social and cognitive outcomes. Table 7.2 provides a detailed list of these contributions.

Context	Contributions
All Studies	Evidence that manipulations in robot gaze can lead to significant social and cognitive outcomes, particularly better information recall, heightened task attentiveness, stronger liking and feelings of groupness, and better task performance led by stronger attributions of mental states.
Study I	**Human Communication** Understanding of the spatial and temporal properties of oratorial speaker gaze behavior speaking American English, particularly those of the speaker's gaze shifts during speech and fixation duration distributions for each gaze target. **Human-Robot Interaction** Evidence that increased robot gaze leads to better information recall and less positive evaluations of the robot in women but does not affect recall or liking in men.
Study II	**Human Communication** Understanding of the spatial and temporal properties of conversational speaker behavior in different participation structures in Japanese, particularly targets, frequencies, and fixation length distributions for gaze shifts toward addressees, bystanders, and overhearers. Understanding of three conversational mechanisms in Japanese: gaze cues that help speakers manage turn exchanges, gaze cues that they use to signal conversational roles, and gaze patterns that signal information structure. **Human-Robot Interaction** Evidence that people follow the norms of the conversational roles that a robot signals to them with high accuracy. Evidence that appropriate signaling of conversational roles can lead to more liking of the robot, feelings of groupness with the robot and others in the conversation, and heightened attentiveness to the conversation.

Study III	**Human Communication** Understanding of how people leak information through gaze cues under cognitive pressure and the temporal and spatial properties of these cues. Evidence that people use information from others' gaze cues—including leakage gaze cues—to make attributions of mental states. **Human-Robot Interaction** Evidence that people read and interpret correctly leakage gaze cues in humanlike robots. Evidence that people's interpretations of leakage gaze cues are affected by the physical design of the robot; they read and correctly interpret leakage gaze cues when produced by a highly humanlike android with little recollection of the presence of these cues, while they did not do so when the cues were produced by a humanlike robot with a stylized, abstract design despite conscious recollection of the presence of these cues. Evidence that people's interpretations of nonverbal cues in robots are affected by whether they own pets; only those who own pets read and interpret correctly leakage gaze cues in humanlike robots. Evidence that robots can also effectively conceal leaked information, but that this behavior negatively affects people's perceptions of the robot's cooperativeness.

Table 7.2. Theoretical contributions of the dissertation.

7.3. Practical Contributions

The practical contributions of this dissertation include a set of design variables for social gaze, a number of data analysis tools for studying social behavior, and computational models of gaze behavior that are created for each empirical study. Table 7.3 provides a detailed list of these contributions.

Context	Contributions
All Studies	A set of design variables for designing social gaze mechanisms. A number of Java-based data analysis tools created for coding and analyzing video, audio, and text data.
Study I	A computational model of oratorial speaker gaze behavior that signals information structure represented as a probabilistic state machine and programmed in C++.
Study II	A computational model of conversational speaker gaze behavior with gaze mechanisms to help manage turn-exchanges, signal conversational roles, and cue information structure represented as a hierarchical probabilistic state machine and programmed in Java.
Study III	A computational model of gaze behavior for producing leakage gaze cues and concealing gaze cues at question-answer sequences represented as a probabilistic state machine and programmed in Java.

Table 7.3. Practical contributions of the dissertation.

7.4. Closing Remarks

Throughout this dissertation, I have argued that humanlike robots can deliver social and cognitive benefits through changes in their social behavior. The three studies that I presented showed that three robotic platforms elicited better information recall, heightened task attentiveness, more liking, stronger feelings of groupness, and stronger attributions of mental states using manipulations in gaze. I have also argued that these benefits can be achieved by following a process of gaining a theoretically and empirically grounded understanding of human communicative processes, carefully designing behavioral mechanisms for humanlike robots that facilitate these processes, and testing how these mechanisms could be manipulated to achieve particular social and cognitive outcomes. In this process, I employed methods and knowledge from a variety of scientific disciplines and made a number of design decisions that were grounded in theory and empirical data. While further work remains in order to improve the validity of these decisions and the generalizability of the results, this dissertation provides a major step towards designing social capabilities for humanlike robots using a theoretically and empirically grounded methodology and understanding their social and cognitive impact in our lives.

Bibliography

Abele, A. (1986). Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior, 10*(2), 83-101.

Ambady, N., & Rosenthal, R. (1992). Thin Slides of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. *Psychological Bulletin, 11*(2), 256–274.

Andersen, H. C. (2001). *Tales, Vol. XVII, Part 3, The Harvard Classics*: P.F. Collier & Son, 1909-14; Bartleby.com.

Argyle, M., Alkema, F., & Gilmour, R. (1971). The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European Journal of Social Psychology, 1*(3), 385-402.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-Contact, Distance and Affiliation. *Sociometry, 28*(3), 289-304.

Argyle, M., & Graham, J. A. (1975). The Central Europe experiment - looking at persons and looking at objects. *Journal of Environmental Psychology and Nonverbal Behaviour, 1*(1), 1-16.

Argyle, M., & Ingham, R. (1972). Gaze, mutual gaze and proximity. *Semiotica, 6*(32-49).

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596-612.

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., & Turk, M. (2005). Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments. *Human Communication Research, 31*(4), 511-537.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators & Virtual Environments, 10*(6), 583-598.

Bailenson, J. N., Yee, N., Patel, K., & Beall, A. C. (2008). Detecting digital chameleons. *Computer in Human Behavior, 24*, 66-87.

Bales, R. F. (1970). *Personality and Interpersonal Behavior*. New York: Holt, Rinehart and Winston.

Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review, 16*(4), 461-468.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41-59). Hillsdale, NJ: Erlbaum.

Bartneck, C., Reichenbach, J., & Breemen, A. (2004). *In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions.* Paper presented at the Design and Emotion Conference (D&E'04), Ankara, Turkey.

Bateson, M., Nettle, D., & Robeerts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology letters, 2*(3), 412.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development, 23*(223-229).

Bayliss, A. P., Pellegrino, G., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology Section A, 58*(4), 631-650.

Bayliss, A. P., & Tipper, S. P. (2006). Predictive Gaze Cues and Personality Judgments: Should Eye Trust You? *Psychological Science, 17*(6), 514-520.

Beattie, G. (1980). The role of language production processes in the organization of behavior in face-to-face interaction. In B. Butterworth (Ed.), *Language Production* (Vol. 1, pp. 69-107).

Becvar, L., Hollan, J., & Hutchins, E. (2008). Representational Gestures as Cognitive Artifacts for Developing Theory in a Scientific Laboratory. In M. S. Ackerman, C. A. Halverson, T. Erickson & W. A. Kellogg (Eds.), *Resources, Co-Evolution, and Artifacts: Theory in CSCW* (pp. 117-143).

Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005). *Towards a humanoid museum guide robot that interacts with multiple persons.* Paper presented at the 5th IEEE-RAS International Conference on Humanoid Robots.

Bente, G., Donaghy, W., & Suwelack, D. (1998). Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. *Journal of Nonverbal Behavior, 22*(1), 31-58.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561-567.

Bouman, C. A. (1997). Cluster: An unsupervised algorithm for modeling Gaussian mixtures. Retrieved from https://engineering.purdue.edu/~bouman/

Breazeal, C. (1998). *A motivational system for regulating human-robot interaction.* Paper presented at the AAAI.

Breazeal, C. (2001). *Emotive qualities in robot speech.* Paper presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'01).

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Breazeal, C., & Scassellati, B. (1999). *How to build robots that make friends and influence people.* Paper presented at the International Conference on Intelligent Robots and Systems (IROS'99).

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

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. (1999). The Cog Project: Building a Humanoid Robot. *Computation for Metaphors, Analogy and Agents, 1562*, 52-87.

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

Bruce, A., Nourbakhsh, I., & Simmons, R. (2002). *The role of expressiveness and attention in human-robot interaction.* Paper presented at the IEEE International Conference on Robotics and Automation (ICRA '02).

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1996). Testing interpersonal deception theory: The language of interpersonal deception. *Communication Theory, 6*, 268-289.

Burnham, T. C., & Hare, B. (2007). Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions in Adult Humans? *Human Nature, 18*(2), 88-108.

Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development, and simulation of everyday mindreading*. Oxford, England: Blackwell.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., et al. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*, 1129-1138.

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI magazine, 22*(4), 67-84.

Cassell, J., Bickmore, T. W., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., et al. (1999a). *Embodiment in Conversational Interfaces: Rea.* Paper presented at the Conference on Human Factors in Computing (CHI'99), Pittsburgh, PA.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K., & Striegnitz, K. (2007). Trading spaces: How humans and humanoids use speech and gesture to give directions *Engineering Approaches to Conversational Informatics*: Wiley.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., et al. (1994). *Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents.* Paper presented at the Computer Graphics Proceedings, Annual Conference Series.

Cassell, J., Torres, O., & Prevost, S. (1999b). Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation *Machine Conversations* (pp. 143-154). Norwell, MA: Kluwer.

Castelli, F., Frith, C. D., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839-1849.

Castiello, U. (2003). Understanding other people's actions: Intention and attention. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 416-430.

Chafe, W. (1979). The flow of thought and the flow of language. *Syntax and semantics, 12*, 159-181.

Chafe, W. (1993). Prosodic and functional units of language. *Talking data: Transcription and coding in discourse research*, 33-43.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893-910.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*(6), 580-601.

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

Clark, H. H., & Carlson, T. B. (1982). Hearers and Speech Acts. *Language, 58*(2), 332-373.

Colburn, A., Cohen, M. F., & Drucker, S. (2000). *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*: Microsoft Research.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior and Development, 23*, 113–118.

Cook, M. (1977). Gaze and Mutual Gaze in Social Encounters. *The American Scientist*, 328-333.

Cook, M., & Smith, J. M. C. (1975). The role of gaze in impression formation. *British Journal of Social and Clinical Psychology, 14*, 19-25.

Coss, R. (1970). The perceptual aspects of eye-spot patterns and their relevance to gaze behaviour. *Behaviour Studies in Psychiatry (Hutt, C. & Hutt, SJ, eds). Pergamon Press, London*, 121—147.

Coutts, L. M., & Schneider, F. W. (1975). Visual behavior in an unfocused interaction as a function of sex and distance. *Journal of Experimental Social Psychology, 11*, 64–77.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). *Wizard of Oz studies: why and how*. Paper presented at the Proceedings of the 1st International Conference on Intelligent User Interfaces.

Dautenhahn, K., Werry, I., Rae, J., & Dickerson, P. (2002). Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot. In K. Dautenhahn (Ed.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Dordrecht: Kluwer Academic Publishers.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin, 129*(1), 74-118.

Duncan, S. (1974). Some signals and rules for taking speaking turns in conversations *Nonverbal Communication: Readings with Commentary*: Oxford University Press.

Duncan, S., Kanki, B., Mokros, H., & Fiske, D. (1984). Pseudounilaterality and other ills to which interaction research is heir. *Journal of Personality and Social Psychology, 46, 1335, 1348*.

Efran, J. S. (1968). Looking for approval: effects on visual behavior of approbation from persons differing in importance. *Journal of Personality and Social Psychology, 10*(1), 21-25.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology, 58*, 342-353.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88-106.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology, 29*, 288-298.

Ellsworth, P. C., Carlsmith, J. M., & Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. *Journal of Social Psychology, 21*, 302-311.

Ellsworth, P. C., & Ludwig, L. M. (1972). Visual behavior in social interaction. *Journal of Communication, 22*, 375-403.

Elzinga, R. H. (1978). Temporal organization of conversation. *Sociolinguistics Newsletter, 9*(2), 21-29.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581-604.

Exline, R., Gray, D., & Schuette, D. (1965a). Visual Behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology, 95*, 201-209.

Exline, R. V. (1963). Explorations in the process of person perception: visual interaction in relation to competition, sex and need for affiliation. *Journal of Personality and Social Psychology, 31*, 1-20.

Exline, R. V., & Eldridge, C. (1967). *Effects of two patterns of a speaker's visual behavior upon the perception of the authenticity of his verbal message.* Paper presented at the Meetings of the Eastern Psychological Association, Boston, MA.

Exline, R. V., Gray, D., & Schuette, D. (1965b). Visual Behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology, 95*, 201-209.

Exline, R. V., & Winters, L. C. (1966). Affective relations and mutual glances in dyads. In S. S. Tomkins & C. Izard (Eds.), *Affect, Cognition, and Personality*. London: Tavinstock.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*, 9602-9605.

Feldman, R. S., Devin-Sheehan, L., & Allen, V. L. (1978). Nonverbal Cues as Indicators of Verbal Dissembling. *American Educational Research Journal, 15*(2), 217-231.

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

Forgas, J. P. (1979). *Social episodes: the study of interaction routines*. London, New York: Published in cooperation with the European Association of Experimental Social Psychology by Academic Press.

Francis, S. J. (1979). Sex differences in nonverbal behavior. *Sex Roles, 5*, 519-535.

Freire, A., Eskritt, M., & Lee, K. (2004). Are Eyes Windows to a Deceiver's Soul? Children's Use of Another's Eye Gaze Cues in a Deceptive Situation. *Developmental Psychology, 40*(6), 1093-1104.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. *Psychological Bulletin, 133*(4), 694-724.

Fry, R., & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. *Journal of Social Psychology, 96*, 145-146.

Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2001). *The impact of eye gaze on communication using humanoid avatars.* Paper presented at the ACM/SIGCHI Conference on Human Factors in Computing Systems (CHI'01).

Garfinkel, H., & Sacks, H. (1970). On formal structures of practical actions. In J. C. McKinney & E. A. Tirayakian (Eds.), *Theoretical sociology* (pp. 338-366). New York: Appleton-Century-Crofts.

Gibson, J. J., & Pick, A. D. (1963). Perception of Another Person's Looking Behavior. *The American Journal of Psychology, 76*(3), 386-394.

Gill, M. J., Swann, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*, 1101-1114.

Gockley, R., Forlizzi, J., & Simmons, R. (2006). *Interactions with a moody robot.* Paper presented at the ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI'06).

Goetz, J., Kiesler, S., & Powers, A. (2003). *Matching robot appearance and behavior to tasks to improve human-robot cooperation.* Paper presented at the 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN'03).

Goffman, E. (1955). On face-work. *Psychiatry, 18*(3), 213-231.

Goffman, E. (1963). *Behavior in Public Places*. Glencoe: The Free Press.

Goffman, E. (1971). *Relations in public*. New York: Harper & Row.

Goffman, E. (1979). Footing. *Semiotica, 25*(1/2), 1-29.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Goodwin, C. (1980). Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry, 50*(272-302).

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Guttman, G., Predovic, M., & Zemanek, M. (1983). *The influence of pet ownership on non-verbal communication and social competence in children.* Paper presented at the International Symposium on the Human-Pet relationship.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*(3), 245-256.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore, MD: Johns Hopkins University Press.

Halliday, M. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Hanks, W. (1996). *Language and Communicative Practices*. Oxford: Westview Press.

Hayduk, L. (1983). Personal space: where we now stand? *Psychological Bulletin, 94*(2), 293-335.

Hemsley, G. D., & Doob, A. N. (1978). The effect of looking behavior on perceptions of a communicator's credibility. *Journal of Applied Social Psychology, 8*, 136-144.

Heylen, D., van Es, I., Nijholt, A., & van Dijk, B. (2005). Controlling the Gaze of Conversational Agents. In J. van Kuppevelt, L. Dybkjaer & O. Bernsen (Eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*: Kluwer Academic Publishers.

Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychological Research, 66*, 174-179.

Hinds, J. (1976). *Aspects of Japanese Discourse Structure*. Tokyo: Kaitakusha.

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Computer Interaction, 19*(1), 151-181.

Hymes, D. (1972). Models of the Interaction of Language and Social Life. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics: The Ethnography of Communication* (pp. 35-77). New York: Holt, Rinehalt & Winston.

Imai, M., Kanda, T., Ono, T., Ishiguro, H., & Mase, K. (2002). *Robot mediated round table: Analysis of the effect of robot's gaze*. Paper presented at the 11th IEEE International Workshop on Robot and Human Interactive Communication.

Ingham, R. J. (1972). *Cross-cultural differences in social behavior.* Oxford University.

Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., & Nakatsu, R. (2001). Robovie: an interactive humanoid robot. *Industrial Robot: An International Journal, 28*(6), 498-504.

Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly, 11*(2), 95-127.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., & Nakatsu, R. (2002). *Development and evaluation of an interactive humanoid robot "Robovie."*. Paper presented at the IEEE International Conference on Robotics and Automation (ICRA'02).

Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., et al. (2007). A humanoid robot that pretends to listen to route guidance from a human. *Autonomous Robots, 22*(1), 87-100.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol (Amst), 26*(1), 22-63.

Kendon, A. (1973). The role of visible behavior in the organization of social interaction. In M. V. Cranach & I. Vine (Eds.), *Social communication and movement: Studies of interaction and expression in man and chimpanzee* (pp. 29-74).

Kendon, A., & Cook, M. (1969). The consistency of gaze patterns in social interaction. *British Journal of Psychology, 69*, 481-494.

Kendon, A., & Ferber, A. (1973). A description of some human greetings. In R. P. Michael & J. H. Cook (Eds.), *Comparative ecology and behaviour of primates* (pp. 591-668).

Kenny, D., & Malloy, T. (1988). Partner effects in social interaction. *Journal of Nonverbal Behavior, 12*(1), 34-57.

Kenny, P., Parsons, T., Gratch, J., Leuski, A., & Rizzo, A. (2007). Virtual Patients for Clinical Therapist Skills Training. *Lecture Notes In Computer Science, 4722*, 197.

Kenny, P., Parsons, T., Gratch, J., & Rizzo, A. (2008). Evaluation of Justina: A Virtual Patient with PTSD. *Lecture Notes In Computer Science, 5208*, 394-408.

Khullar, S., & Badler, N. (2001). Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems, 4*(1), 9-23.

Kiesler, S. (2005). *Fostering common ground in human-robot interaction.* Paper presented at the IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN'05).

Kiesler, S., & Goetz, J. (2002). *Mental models of robotic assistants.* Paper presented at the ACM/SIGCHI Conference on Human Factors in Computing Systems (CHI'02).

Kleck, R. E., & Nuessle, W. (1968). Congruence between the indicative and communicative functions of eye-contact in interpersonal relations. *British Journal of Social and Clinical Psychology, 7*, 241-246.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 78-100.

Krauss, R., Chen, Y., & Chawla, P. (1996). Nonverbal Behavior and Nonverbal Communication: What Do Conversational Hand Gestures Tell Us? *Psychological Science, 7*(4), 226-231.

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology, 40*, 312-320.

Kraut, R. E., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39*, 784-798.

Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A., & Kuzuoka, H. (2007). *Museum guide robot based on sociological interaction analysis.* Paper presented at the ACM/SIGCHI conference on Human Factors in Computing Systems (CHI'07).

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology, 36*, 635-642.

Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*, 541-567.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4*(2), 50-59.

Lee, S., Lau, I. Y., Kiesler, S., & Chiu, C. Y. (2005). *Human Mental Models of Humanoid Robots*. Paper presented at the 2005 IEEE International Conference on Robotics and Automation (ICRA'05).

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412-426.

Libby, W. L. (1970). Eye contact and direction of looking as stable individual differences. *Journal of Experimental Research in Personality, 4*, 303-312.

Lutchmaya, S., & Baron-Cohen, S. (2002). Human sex differences in social and non-social looking preferences, at 12 months of age. *Infant Behavior and Development, 25*, 319-325.

Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002). Foetal testosterone and eye contact in 12-month-old human infants. *Infant Behavior and Development, 25*, 327-335.

Malle, B., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*(2), 101-121.

Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. *Psychological Science, 16*, 236-239.

Matsusaka, Y., Fujie, S., & Kobayashi, T. (2001). *Modeling of Conversational Strategy for the Robot Participating in the Group Conversation.* Paper presented at the Seventh European Conference on Speech Communication and Technology (ISCA'01).

Maynard, S. K. (1989). *Japanese Conversation: Self-Contextualization through Structure and Interactional Management*. Norwood, NJ: Alex.

McLaughlin, M., & Cody, J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Human Communication Research, 8*, 299-316.

Miklósi, A., Polgárdi, R., Topál, J., & Csányi, V. (2000). Intentional behavior in dog-human communication: An experimental analysis of "showing" behaviour in the dog. *Animal Cognition, 3*(3), 159-166.

Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android robot for studying human-robot interaction. *Lecture Notes In Computer Science: Proceedings of the 17 th international conference on Innovations in applied artificial intelligence, 17*(20), 424-434.

Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., et al. (2004). *Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII.* Paper presented at the IEEE Conference Intelligent Robots and Systems (IROS'04).

Mobbs, N. A. (1968). Eye-contact in relation to social introversion-extroversion. *British Journal of Social and Clinical Psychology, 7*, 305-306.

Moon, Y., & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651-674.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human-Computer Studies, 49*, 79-94.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction, 14*, 395-435.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. *Journal of Applied Social Psychology, 29*(5), 1093-1109.

Nass, C., Moon, Y., & Green, T. (1997). Are computers gender neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*(10), 864-876.

Nass, C., & Steuer, J. (1993). Computers and Social Actors. *Human Communication Research, 19*(4), 504-527.

Nichols, K., & Champness, B. (1971). Eye gaze and the GSR. *Journal of Experimental Social Psychology, 7*, 623-626.

Nielsen, G. (1962). *Studies in Self Confrontation*. Copenhagen: Monksgaard.

Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: Teleoperated Android of an Existing Person. In A. C. de Pina Filho (Ed.), *Humanoid Robots: New Developments*. Vienna, Austria: I-Tech Education and Publishing.

Noller, P. (1984). *Nonverbal Communication and Marital Interaction*: Pergamon.

Nourbakhsh, I. (1999). An affective mobile robot educator with a full-time job. *Artificial Intelligence, 114*(1-2), 95-124.

Nowak, K. (2001). *Defining and differentiating copresence, social presence and presence as transportation*. Paper presented at the Presence Conference.

Nowak, K. L., & Biocca, F. (2003). The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators \& Virtual Environments, 12*(5), 481-494.

Olafsen, K. S., Ronning, J. A., Kaaresen, P. I., Ulvund, S. E., Handegard, B. H., & Dahl, L. B. (2006). Joint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention. *Infant Behavior and Development, 29*(4), 554-563.

Otteson, J. P., & Otteson, C. R. (1980). Effects of teacher gaze on children's story recall. *Perceptual and Motor Skills, 50*, 35-42.

Ozaki, Y. T. (1970). *The Japanese Fairy Book*. Tokyo: Tuttle Publishing.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1998). *My partner is a real dog: cooperation with social agents*. Paper presented at the 1996 ACM conference on Computer supported cooperative work.

Patterson, M. L. (1976). An arousal model of interpersonal intimacy. *Psychological Review, 83*, 235-245.

Patterson, M. L., Webb, A., & Schwartz, W. (2002). Passing encounters: atterns of recognition and avoidance in pedestrians. *Basic and Applied Social Psychology, 24*(1), 57-66.

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: Processing of mutual and averted gaze in the superior temporal sulcus. *Psychological Science, 15*, 598-603.

Perrett, D. I., & Emery, N. J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. *Current Psychology of Cognition, 13*(5), 683-694.

Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J., & Rolls, E. T. (1992). Organisation and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B, 335*, 23-30.

Peters, C. (2005). Direction of attention perception for conversation initiation in virtual environments. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier & T. Rist (Eds.), *Lecture notes in computer science* (pp. 215-228). London: Springer-Verlag.

Peters, C., & O'Sullivan, C. (2003). *Bottom-up visual attention for virtual human animation.* Paper presented at the 16th International Conference on Computer Animation and Social Agents.

Pierno, A. C., Becchio, C., Wall, M. B., Smith, A. T., Turella, L., & Castiello, U. (2006). When gaze turns into grasp. *Journal of Cognitive Neuroscience, 18*, 2130-2137.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems, 42*(3-4), 271-281.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology Section A, 32*(1), 3-25.

Powers, A., & Kiesler, S. (2006). *The advisor robot: tracing people's mental model from a robot's physical attributes*. Paper presented at the Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction.

Prince, E. F. (1981). Toward a taxonomy of given-new information. *Radical Pragmatics*, 223—255.

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

Rehm, M., & Andre, E. (2005). *Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent.* Paper presented at the International Conference on Intelligent Virtual Agents (IVA'05).

Rickel, J., & Johnson, W. L. (1999). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence, 13*, 343-382.

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy sciences, 4*(2), 155-169.

Roberts, C. (1996). *Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics*: Ohio State University, Department of Linguistics.

Rowe, K. J., & Rowe, K. S. (1999). Investigating the relationship between students' attentive-inattentive behaviours in the classroom and their literacy progress. *International Journal of Educational Research, 31*(1-2), 81-117.

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning, 19*(2), 195-208.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K., et al. (2002). *The intelligent ASIMO: system overview and integration*, IEEE/RSJ International Conference on Intelligent Robots and System (IROS'02).

Scassellati, B. (2001). *Foundations for a theory of mind for a humanoid robot.* MIT, Cambridge, MA.

Scheeff, M., Pinto, J., Rahardja, K., Snibbe, S., & Tow, R. (2000). *Experiences with Sparky: A social robot.* Paper presented at the Workshop on Interactive Robot Entertainment.

Schegloff, E. A. (1968). Sequencing in Conversational Openings. *American Anthropologist, 70*(6), 1075-1095.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society, 29*(1), 1-63.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotion from content-masked speech. *Journal of Psycholinguistic Research, 1*, 269-285.

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7*, 31-44.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*: Basic Books.

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition, 12*, 127-144.

Sherwood, J. (1987). Facilitative effects of gaze upon learning. *Perceptual and Motor Skills, 2/3*(2), 1275-1278.

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). *Where to look: a study of human-robot engagement*. Paper presented at the 9th international Conference on Intelligent User Interfaces (IUI'04), Funchal, Madeira, Portugal.

Simmel, G. (1921). Sociology of the senses: Visual interaction. In R. E. Park & E. W. Burgess (Eds.), *Introduction to the Science of Sociology*: University of Chicago Press.

Simon, H. (1996). *The sciences of the artificial*: MIT press.

Snyder, M., Grather, J., & Keller, K. (1974). Staring and Compliance: A Field Experiment on Hitchhiking. *Journal of Applied Social Psychology, 4*(2), 165-170.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*(2), 97-124.

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design, 2*(1), 55-65.

Strongman, K. T., & Champness, B. G. (1968). Dominance hierarchies and conflict in eye contact. *Acta Psychologica, 28*, 376-386.

Surakka, V., & Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology, 29*, 23-33.

Tannen, D. (1984). *Conversational Style: Analyzing Talk among Friends*. Norwood, NJ: Ablex.

Thomas, F., & Johnston, O. (1981). *Disney animation: The illusion of life*: Abbeville Press New York.

Thorisson, K. R. (2002). Natural turn-taking needs no manual: Computational theory and model, from perception to action. In B. Granstrom, D. House & I. Karlsson (Eds.), *Multimodality in Language and Speech Systems* (pp. 173-207): Kluwer Academic Publishers.

Tojo, T., Matsusaka, Y., Ishii, T., & Kobayashi, T. (2000). *A conversational robot utilizing facial and body expressions*. Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics.

Trafton, J. G., Bugajska, M. D., Fransen, B. R., & Ratwani, R. M. (2008). *Integrating vision and audition withing a cognitive architecture to track conversations*. Paper presented at the 3rd ACM/IEEE international conference on Human Robot Interaction (HRI'08).

Van Houten, R., Nau, P., Mackenzie-Keating, S., Sameoto, D., & Colavecchia, B. (1982). An analysis of some variables influencing the effectiveness of reprimands. *Journal of Applied Behavior Analysis, 15*, 65-83.

Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). *Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes*. Paper presented at the ACM/SIGCHI conference on Human factors in computing systems (CHI'01).

Vertegaal, R., van der Veer, G., & Vons, H. (2000). *Effects of Gaze on Multiparty Mediated Communication*. Paper presented at the Graphics Interface.

Vilhjalmsson, H. H., & Cassell, J. (1998). *BodyChat: autonomous communicative behaviors in avatars*. Paper presented at the 2nd international Conference on Autonomous Agents Minneapolis, MN.

Von Cranach, M., & Ellgring, J. H. (1973). The perception of looking behaviour. In M. Von Cranach & I. Vine (Eds.), *Social Communication and Movement*. London: Academic Press.

Wang, N., Johnson, W. L., Rizzo, P., Shaw, E., & Mayer, R. E. (2005). *Experimental evaluation of polite interaction tactics for pedagogical agents*. Paper presented at the 10th International Conference on Intelligent User Interfaces (IUI'05).

Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics, 32*, 1177—1207.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive an negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Watson, O. M. (1970). *Proxemic behavior: a cross-cultural study*. The Hague: Mouton.

Waxer, P. H. (1977). Nonverbal cues for anxiety: an examination of emotional leakage. *Journal of Abnormal Psychology, 86*(3), 306-314.

Weisbrod, R. M. (1965). Looking Behavior in a Discussion Group. Unpublished Paper. Department of Psychology, Cornell University.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 13*(2), 183-194.

Williams, K. (2001). *Ostracism: The power of silence*: The Guilford Press.

Williams, K., Cheung, K. T., & Choi, W. (2000). Cyberostracisms: Effects of being ignored over the internet. *Journal of Personality and Social Psychology, 79*, 748-762.

Williams, L. M., Senior, C., David, A. S., Loughland, C. M., & Gordon, E. (2001). In Search of the" Duchenne Smile": Evidence from Eye Movements. *Journal of Psychophysiology, 15*(2), 122-127.

Wolf, T., Rode, J., Sussman, J., & Kellogg, W. (2006). *Dispelling" design" as the black art of CHI*.

Wolff, P. H. (1963). Observations on the early development of smiling. In B. M. Foss (Ed.), *Determinants of Infant Behavior* (Vol. 2, pp. 113-138). London: Methuen.

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

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.* Paper presented at the ACM/SIGCHI Conference on Human Factors in Computing Systems (CHI'08), Florence, Italy.

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., & Miyamoto, T. (2006). *Responsive robot gaze to interaction partner.* Paper presented at the Robotics: Science and Systems.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-59). New York: Academic Press.

Appendix A

Human Storyteller Gaze Length Distribution Parameters in Study I

This appendix includes the gaze length parameters for each cluster identified from empirical results. The top two rows show how frequently these clusters were looked at and the total time the storyteller spent looking at these clusters. The two middle rows are the *mean* and *standard distribution* parameters used to generate gaze lengths for ASIMO over a Normal distribution. As discussed in Section 3.5, a more careful post-hoc analysis showed that these gaze lengths can be better modeled using a two-parameter continuous distribution such as the Gamma distribution. The bottom two rows provide the *shape* and *scale* parameters for Gamma distributions fitted to the data from each cluster.

	Listener 1	*Listener 2*	*Fixed spot*	*Environment*
Frequency (%)	13	11	38	38
Time spent (%)	38	27	30	5
Mean (seconds)	2.64	2.26	2.64	1.07
StDev (seconds)	1.89	1.24	2.48	0.92
Shape (k)	3.32	2.72	1.38	2.19
Scale (θ)	0.68	0.97	1.92	0.49

Table A.1. Gaze length distribution parameters for the four gaze clusters identified in the first study.

Appendix B

Gaze Algorithm Designed for ASIMO in Study I

This appendix presents the gaze algorithm suggested by Cassell et al. (1999b) and the algorithm that directed ASIMO's gaze in the first study. Cassell et al. (1999b) suggested the following algorithm to simulate natural gaze behavior using a randomized function, *distribution(x)*, that returns true with probability *x*.

```
for each proposition do
    if proposition is theme then
        if beginning of turn or distribution(0.70) then
            attach a look-away from the listener
        end if
    else if proposition is rheme then
        if end of turn or distribution(0.73) then
            attach a look-toward the listener
        end if
    end if
end for
```

In the designed algorithm, *distribution(x)* produces a uniform randomized function that returns true with the probability derived from the algorithm described by Cassell et al. (1999b) and from the empirical data for each gaze cluster (listener 1, listener 2, fixed spot, and environment). Function *length(x)* generates a duration for the gaze over a normal distribution with mean and standard deviation values from the empirical results (~Normal(Mean(x), StDev(x))). Below is the designed algorithm.

```
for each part of the utterance (theme/rheme/pause) do
    while the duration of the part do
        if current part is pause then
            if distribution(probability(environment)) then
                gaze at environment with length(environment)
```

```
            else
                gaze at fixed spot with length(fixed spot)
            end if
        else if current part is theme then
            if distribution(0.70) then
                if distribution(probability(environment) then
                    gaze at environment with length(environment)
                else
                    gaze at fixed spot with length(fixed spot)
                end if
            else
                if distribution(probability(listener 1)) then
                    gaze at listener 1 with length(listener 1)
                else
                    gaze at listener 2 with length(listener 2)
                end if
            end if
        else if current part is rheme then
            if distribution(0.73) then
                if distribution(probability(listener 1)) then
                    gaze at listener 1 with length(listener 1)
                else
                    gaze at listener 2 with length(listener 2)
                end if
            else
                if distribution(probability (environment)) then
                    gaze at environment with length(environment)
                else
                    gaze at fixed spot with length(fixed spot)
                end if
            end if
        end if
    end while
end for
```

Appendix C

Gaze Length Distributions in Study II

This appendix presents the *means*, *standard deviations*, and *shape* and *scale* parameters for the fitted Gamma distributions for gaze lengths for each target cluster in the three studied conversational structures: two-party conversation, two-party-with-bystander conversation, and three-party conversation.

	Time spent (%)	Frequency (%)	Mean (seconds)	StDev (seconds)	Shape (k)	Scale (θ)
Two-party conversation						
Addressee's face	26	33	0.95	0.91	1.65	0.56
Addressee's body	48	36	0.99	1.03	1.92	0.84
Environment	26	31	1.01	0.98	0.90	1.14
Two-party-with-bystander conversation						
Addressee's face	25	31	1.40	1.30	0.74	1.55
Addressee's body	51	36	1.01	1.22	1.72	1.20
Bystander's face	5	7	0.77	0.58	2.19	0.44
Bystander's body	3	5	0.71	0.49	1.76	0.57
Environment	16	21	0.96	1.04	1.84	0.59
Three-party conversation						
First addressee's face	21	16	0.98	1.26	1.25	1.26
First addressee's body	7	8	0.80	0.83	1.61	0.62
Second addressee's face	35	26	0.97	0.83	1.71	0.93
Second addressee's body	8	10	0.85	0.79	2.23	0.41
Environment	29	29	0.82	0.78	1.28	0.70

Table C.1. Gaze length distribution parameters for all targets in three conversational structures.

Appendix D

Gaze Patterns that Signal Japanese Information Structure in Study II

The modeling of the relationship between gaze cues and information structure of the spoken discourse, as described in Section 4.1.2.2, identified a number of recurring patterns of gaze shifts that are initiated at the onset of thematic segments. The analysis identified two recurrent patterns in the two-party and two-party with bystander conversations and another set of two patterns in the three-party conversation. This appendix provides graphical representations of these patterns and a table that lists the frequencies of occurrence for each pattern.

Patterns Identified in Two-party/Two-party-with-bystander Conversations

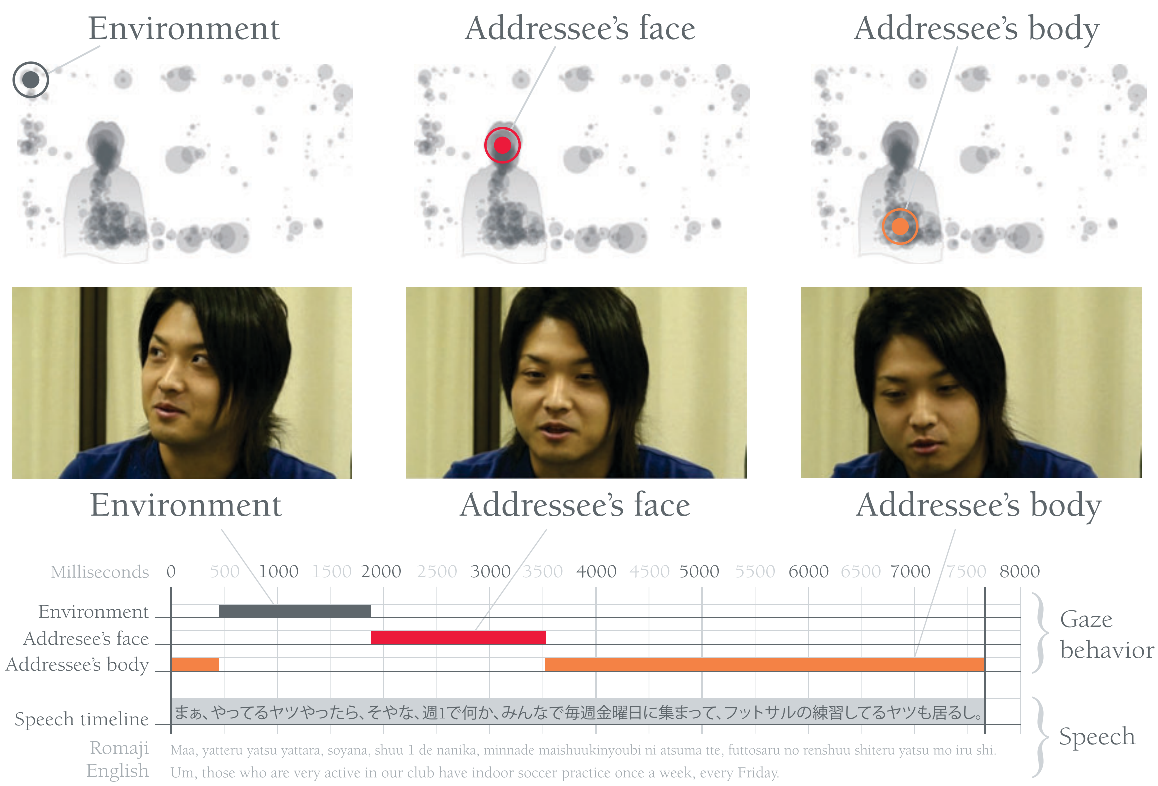


Figure D.1. The most frequent pattern (63% of the time) observed at turn beginnings and the second most frequent (25% of the time) pattern observed at thematic field beginnings in the two-party/two-party-with-bystander conversations.

Figure D.2. The second frequent pattern (17% of the time) observed at turn beginnings and the most frequent (30% of the time) pattern observed at thematic field beginnings in the two-party/two-party-with-bystander conversations.

Patterns Identified in Three-party Conversations

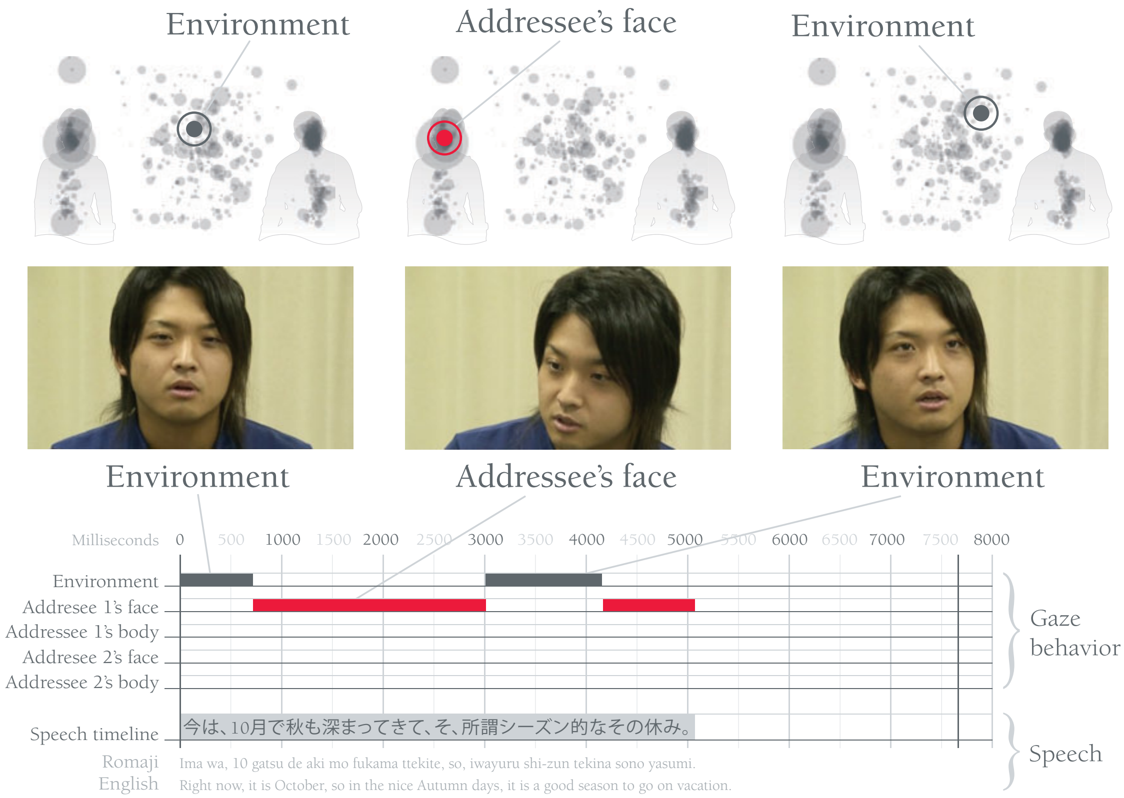


Figure D.3. The most frequent pattern (60% of the time) observed at turn beginnings and the most frequent (47% of the time) pattern observed at thematic field beginnings in the three-party conversations.

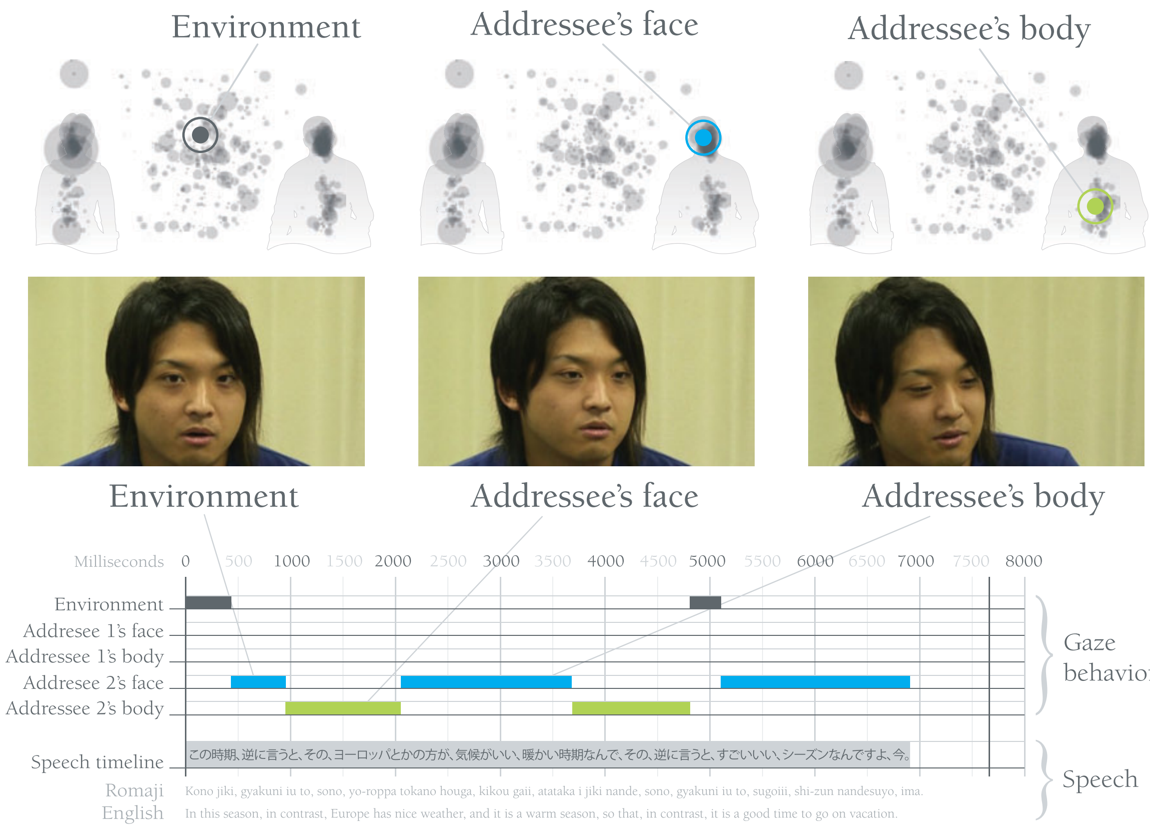


Figure D.4. The second most frequent pattern (7% of the time) observed at turn beginnings and the second most frequent (29% of the time) pattern observed at thematic field beginnings in the three-party conversations.

	Two-party conversations	Three-party conversation
Look away > Look at > Look down	***25%*** at thematic field beginnings ***63%*** at turn beginnings	***29%*** at thematic field beginnings ***7%*** at turn beginnings
Look at > Look down > Look at	***30%*** at thematic field beginnings ***17%*** at turn beginnings	*Not observed*
Look away > Look at > Look away	*Not observed*	***47%*** at thematic field beginnings ***60%*** at turn beginnings
Pattern continuing from the previous thematic field	***22%*** at thematic field beginnings ***0%*** at turn beginnings	***22%*** at thematic field beginnings ***0%*** at turn beginnings
No recurring pattern	***22%*** at thematic field beginnings ***21%*** at turn beginnings	***2%*** at thematic field beginnings ***33%*** at turn beginnings

Table D.1. Frequencies of the patterns identified in the two- and three-party conversations. Frequencies from two-party and two-party-with-bystander conversations are combined because similar patterns with similar frequencies were observed in these two conversations.

Appendix E

Summary of Design Elements for Robovie's Gaze Behavior in Study II

	Two-party conversation	Two-party-with-bystander conversation	Three-party conversation
Greeting	Acknowledge the addressee	Acknowledge the addressee and then the bystander	Acknowledge one of the addressees and then the other addressee
Participant structure (footing)	Direct gaze at the addressee at the transition from greeting to casual conversation and keep direction of gaze at the addressee at all times	Direct attention at the addressee at the transition from greeting to casual conversation and keep direction of attention mostly at the addressee occasionally glancing at the bystander for short periods	Divide attention at both addressees at the transition from greeting to casual conversation producing turn-yielding signals for both addressees and wait for one of the them to take the floor Switch speakers at "paragraphs"
Conversation structure (turn-exchanges)	*Turn-yielding:* Look at the addressee at the end of a turn *Turn-taking:* Look at the addressee during minimal responses and look away from the addressee at the beginning of the turn	*Turn-yielding:* Look at the addressee at the end of a turn *Turn-taking:* Look at the addressee during minimal responses and look away from the addressee at the beginning of the turn	*Turn-yielding:* Look at the one of the addressees at the end of a turn *Turn-yielding with speaker change:* Look at one of the addressees and then the other and wait for one of them to take the floor *Turn-taking:* Look at the addressee who just passed the floor during minimal responses and look away at the beginning of the turn

Information structure (thematic fields)	Look in pattern "*Look away > Look at > Look down*" at the addressee Look in pattern "*Look down > Look at > Look down*" at the addressee	Look in pattern "*Look away > Look at > Look down*" at the addressee Look in pattern "*Look down > Look at > Look down*" at the addressee Short glances at the bystander at random intervals	Look in pattern "*Look away > Look at > Look away*" at one addressee at a time but at both addressees Look in pattern "*Look away > Look at > Look down*" at one addressee at a time but at both addressees
Leave-taking	Acknowledge the addressee	Acknowledge the addressee and then the bystander	Acknowledge one of the addressees and then the other addressee

Table E.1. A summary of the gaze mechanisms designed for Robovie in Study II.

Appendix F

Length Distributions for Leakage and Concealing Gaze Cues in Study III

This appendix illustrates human and robot fitted Normal distributions of leakage and concealing gaze cue lengths. Cue lengths for the robots were calculated modifying the gaze length distributions from the human data to optimize for the motor capabilities of the two robots for smooth and natural motion and keeping the total gaze durations for the two robots equal. Overall, robots' gaze cues were designed to be longer with smaller variance due to the base delays added to the distributions obtained from the human data.

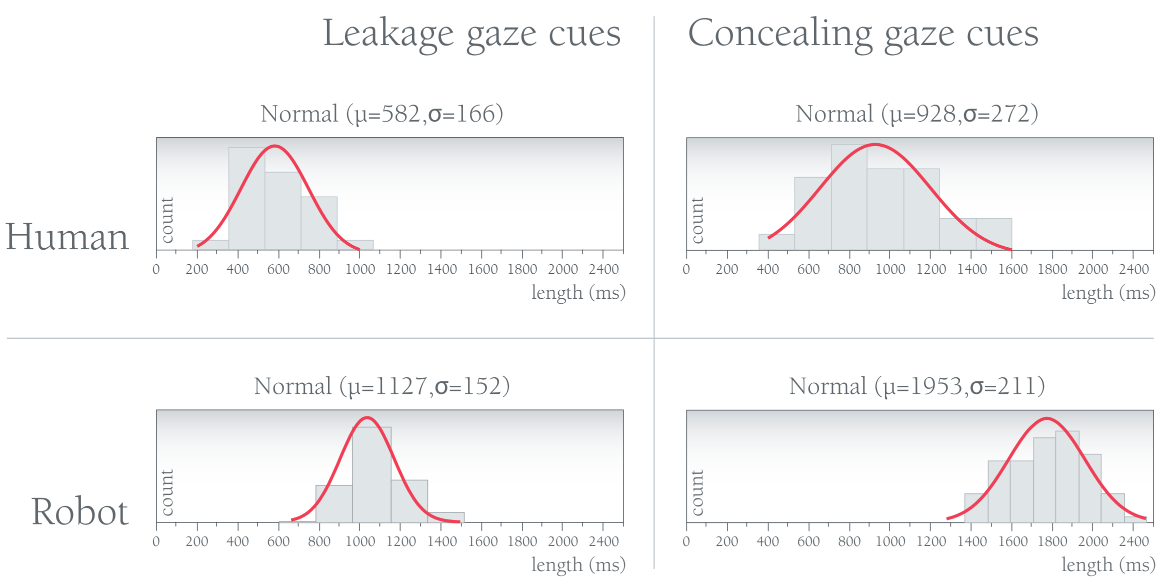


Figure F.1. Leakage and concealing gaze cue length distributions calculated from the human data and those created for the robots.