Nonverbal Leakage in Robots: Communication of Intentions through Seemingly Unintentional Behavior

Bilge Mutlu[1], Fumitaka Yamaoka[2,3], Takayuki Kanda[2], Hiroshi Ishiguro[2,3], Norihiro Hagita[2]

(1) Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, USA
bilge@cs.cmu.edu

(2) ATR Intelligent Robotics and Communication Laboratory
2-2-2 Hikaridai, Keihanna, Kyoto, Japan
{yamaoka, kanda, hagita}@atr.jp

(3) Faculty of Engineering
Osaka University
2-1 Yamadaoka, Suita City, Osaka, Japan
ishiguro@ams.eng.osaka-u.ac.jp

ABSTRACT

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 on the emotional state and intentions of an individual. One family of such cues is called "nonverbal leakage." In this paper, we explore whether people can read nonverbal leakage cues—particularly gaze cues—in humanlike robots and make inferences on robots' intentions, and whether the physical design of the robot affects these inferences. We designed a gaze cue for Geminoid—a highly humanlike android—and Robovie—a robot with stylized, abstract humanlike features—that allowed the robots to "leak" information on what they might have in mind. In a controlled laboratory experiment, we asked participants to play a game of guessing with either of the robots and evaluated how the gaze cue affected participants' task performance. We found that the gaze cue did, in fact, lead to better performance, from which we infer that the cue led to attributions of mental states and intentionality. Our results have implications for robot design, particularly for designing expression of intentionality, and for our understanding of how people respond to human social cues when they are enacted by robots.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – *Human factors*. H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology, User-Centered Design*.

General Terms

Design, Human Factors

Keywords

Nonverbal Behavior, Gaze, Nonverbal Leakage, Humanlikeness, Robovie, Geminoid

1. INTRODUCTION

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 understand 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 his or her emotional or intellectual state. These examples illustrate a set of behaviors called "nonverbal leakage" cues that are

HRI'09, March 11–13, 2009, La Jolla, California, USA.

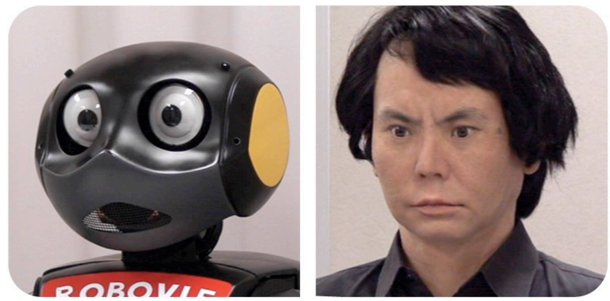

Figure 1. The humanlike robots we used in our study, Robovie R-2 (left) and Geminoid (right).

products of internal, cognitive processes and reveal to others information on the emotional state or intentions of an individual [19,58].

Research in human communication has shown that, using nonverbal cues, naïve observers can identify deception [18,19], dissembling [22], genuineness of smiles [52,57], friendliness and hostility [2], affective states [36,48,49,56], and disfluency of speech [17]. Furthermore, these behaviors might play an important role in forming impressions of others—a process in which people rely heavily on nonverbal behavior [1]. We argue that nonverbal leakage cues and, more broadly, seemingly unintentional and non-semantic nonverbal behaviors pose an important area of inquiry for 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 explicit, deliberate cues such as communication of primary emotions through facial expressions [5,8,10,29,40,47], arm and bodily gestures [53], and vocal tone [9] and communication of attention through gaze [11,27,33,41,50,54] and pointing gestures [28,51], how implicit, non-strategic, and non-semantic cues might be used in human-robot communication has not been explored.

Do people detect nonverbal leakages in robots? If so, do they interpret these messages correctly to attribute intentions to the robot? How do the physical characteristics of the robot affect these inferences? In this paper, we attempt to answer these questions focusing on gaze cues, which are found to be a particularly salient set of nonverbal cues in the communication of complex mental states and intentionality [4,21,24]. Our study follows a process of gaining a better understanding of the concept of leakage from theory on nonverbal communication and observations of human behavior, designing behavioral cues for humanlike robots (as, in this study, those shown in Figure 1), contextualizing these behaviors in human-robot interaction scenarios, and evaluating whether these cues communicate intentions and states of mind by testing theoretically based hypotheses in an empirical study. In the remainder of this paper we summarize related work on non-semantic cues, nonverbal leakage, and gaze cueing, describe our methodology, present our results, and discuss implications of our findings for human-robot interaction research.

2. RELATED WORK

In this section, we summarize related work on nonverbal leakage and focus particularly on how people interpret gaze cues and attribute mental states and intentions to others. Next, we provide a brief summary of existing research on nonverbal leakage in human-robot interaction and embodied virtual agents. We also provide background on how the humanlikeness of a robot might affect perceptions of its nonverbal cues.

Nonverbal Leakage Cues

Imagine two friends, Akira and Mai playing the popular guessing game "two truths and a lie." Akira will tell three facts about himself, of which two are true and one is a lie, and Mai will try to guess which one of the facts is a lie. In guessing Akira's lie, Mai will primarily rely on her knowledge of Akira's background, her world experience, and skills such as empathy. However, she will also look for signs of apprehension, guilt, or excitement in Akira's nonverbal and vocal behavior, and in the presence of such signs, use them to infer which one of Akira's facts is a lie. In this scenario, the signs that Mai is looking for are behaviors that Akira will unintentionally produce due to heightened arousal, his own feelings such as guilt, attempts to control his behaviors and feelings, and/or the cognitive complexity required to manufacture the lie [12,18,58]. Mai, on the other hand, will show an automatic and unconscious propensity to search for and respond to these signals [52,57].

The scenario above illustrates a common process in interpersonal communication in which people use unintentionally produced, non-strategic, and non-semantic nonverbal cues to infer the intentions or the emotional and intellectual states of communication partners. A particular type of such processes is "nonverbal leakage"—as termed by Ekman and Freisen [19]—in which feelings or thoughts "leak" through the nonverbal channel to reveal the internal state of an individual. 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 [19].

Research has also shown that people automatically search for cues that might leak information in others' nonverbal behaviors. 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 [57], to distinguish genuine smiles—called the "Duchenne Smile" [20]—from smiles of appeasement [52]. Furthermore, identifying these cues and interpreting their meanings can be done by naïve observers with no particular expertise. 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 [56]. Similarly, Feldman and his colleagues [22] 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 [17] 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 [2].

Gaze Cues as a Channel of Nonverbal Leakage

Gaze cues are a particularly important set of leakage cues that provide a wealth of information on the mental states and intentions of an individual [4,7,21,23,24,25,35,37,45]. Social and developmental psychological studies have shown that through observing others' gaze patterns, people infer personality traits [35]—particularly trustworthiness [6]—and detect and infer deception [4,25,37,38]. For instance, Freire and his colleagues [23] 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 and intentions based on information from these cues [4,7,21,45]. Emery [21] 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 [4] 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 his 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 [13,14]. Similarly, research has also found behavioral evidence that people's motor intentions can be inferred by monitoring their gaze direction [15,46].

Leakage Cues in Human-Robot Interaction

Research in human-robot interaction has focused mainly on creating explicit expressions of emotional states and intentions [5,8,9,11,27-29,33,40,41,47,50,51,54] and has not looked at whether these states and intentions could be communicated through implicit, seemingly unintentional cues with the exception of a single study on virtual agents. Bailenson and his colleagues [3] asked participants to interact with an agent that mimicked participants' nonverbal behavior—a common unconscious behavior seen in human communication called the "chameleon effect" [16]—and to rate whether they thought that the agent was a human or a computer (as in a "Turing Test"). They found that participants (failingly) 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.

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 [42,43]. People reliably rate agents with highly humanlike features to be more socially attractive, more satisfactory as partners [42], more co-present [43], and more likeable [44] than agents with less humanlikeness and cooperate more with them [44]. Research in human-robot interaction has shown similar attributions to robots [26,30,34]. Hinds and her colleagues [26] 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 and her colleagues [30] found that people expected the physical and behavioral

Figure 2. The spatial configuration of the game task.

characteristics of a robot to match its task and complied more with instructions given by a robot that met their expectations of appropriateness. This expectation of appropriateness in a robot's appearance and behavior suggests that people might expect and correctly interpret leakage cues when they are enacted by robots with the appropriate level of humanlikeness. In support of this hypothesis, Kiesler and Goetz [34] argued that humanlike characteristics might engender a more human mental model of a robot. In the next section, we describe the methodology of our investigation.

3. METHODOLOGY

To gain a better understanding of how gaze cues might serve to communicate mental states and intentions, we choreographed a dyadic game task and observed whether human players "leaked" information through their gaze behavior and, if so, how. We used the findings from our observation to design gaze cues for two robots: Geminoid and Robovie R-2 (see Figure 1). In a controlled experiment, we asked naïve participants to play the game with either one of the two robots and measured how the gaze cue affected participants' performance in the game. In the following paragraphs, we describe our interaction design of the game task , findings from observations of human players, design of the gaze cues for the robots, and evaluation of the designed gaze cues, including experimental design, hypotheses, study procedure, measurements, and participant profile.

3.1 Interaction Design of the Experiment

We devised an experimental task 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," chose an item—without identifying it to the other player—among fourteen items placed on a table located between the two players (see Figure 2). The other player, the "guesser," tried to guess which item the picker chose by asking the picker a set of questions that can be answered with "Yes" and "No." We carefully chose the items on the table from artifacts that are commonly used in Japanese daily life and that represent a balanced set of colors, shapes, materials, and sizes. We placed the items on the table equadistantly and determined their spread so that the players did not have to move their heads to glance at the items.

We provided participants 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. We empirically determined the number of items on the table to be fourteen in order to allow participants to identify the item with an expected average of five questions.

Observations of Leakage Cues in Human-Human Interaction
To understand whether, and if so, how, human pickers would leak information on their choice of items, we hired two all-male dyads and asked members of the dyads to play the game. Each participant played the roles of picker and guesser. We captured pickers' gaze behavior using high-definition cameras and conducted a frame-by-frame analysis of the video sequences. The most significant finding of our analysis was that pickers often looked towards their pick immediately before answering questions in very short glances, verifying that they know the answer to the question while trying to conceal the behavior. The top row in Figure 3 shows image sequences of one of these glances.

The Design of the Leakage Cue for the Robots
We used the main finding of our analysis directly to design gaze cues for the robots. In our design, the robots produced two 400-millisecond glances at the object that it picked immediately before answering two of the first three questions that the participant directed at the robot. Image sequences of both robots' enactments of the gaze cue are shown in the middle and bottom rows of Figure 3. The glances at the objects took a total of 1200 milliseconds including travel time of the eyes to and from the gaze target. We determined the gaze duration and timing optimizing for the motor capabilities of the two robots for smooth and natural motion while considering the range of the lengths of glances that we found in our analysis of the data from human participants.

In interacting with the participants, the robots followed common interaction rituals [31]. They introduced themselves to the participants, provided them with information on the task, maintained fluency in the interaction using phrases such as "Let's play one more time," and ended the interaction appropriately by thanking the participant for playing the game. We used pre-recorded human voice to create a rich library of utterances for the robots. Each expression was recorded several times in different forms and inflections. In producing these expressions, the robots randomly chose from a set of alternatives. In performing the task, the robots did not use speech recognition. Instead, a human operator initiated the robots' speech by selecting expressions from a library.

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. In designing Geminoid's gaze behavior, we added random eye blinks

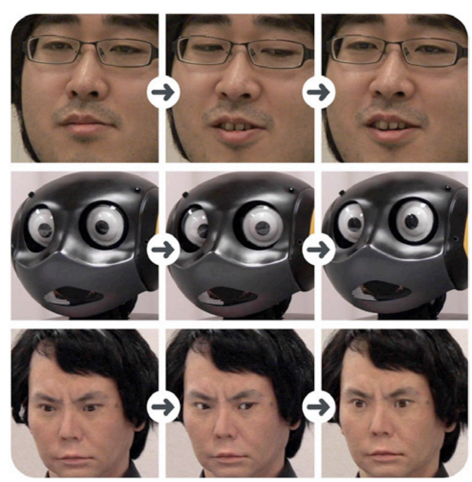

Figure 3. Image sequences of the leakage gaze cue from our observation of human players (top row) and of our design of the cue for Robovie (middle row) and Geminoid (bottom row).

at an average interval of five seconds. We also produced lip movements synchronized with the robot's speech by capturing our voice actor's lip movements using a five-camera motion-capture system. Finally, we differentiated Geminoid's voice 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. We used post-processing to differentiate the voices in order to maintain the same length and inflections for each expression between the two robots.

3.2 Experimental Design

We conducted a two-by-two (two robots and "gaze cue" vs. "no gaze cue" conditions), mixed-factorial-design experiment 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. In half of these trials (excluding the practice trial), the robot produced the gaze cue, glancing at its pick (as illustrated in the middle and bottom rows of Figure 3). We delayed the robot's answers before which it did not produce the gaze cue with the duration of the glance to keep the time it took the robot to answer questions consistent across trials and conditions. In summary, the two gaze conditions were as follows:

In **condition 1**, after the question, the robots waited (the same amount of time that a glance took), looked up, establishing eye contact, and answered the participant's question.

In **condition 2**, after the question, the robots glanced at the object, looked up, establishing eye contact, and answered the participant's question.

Except the two short glances, the robot's behaviors were identical across trials. Each participant played the game four times in each condition with either one of the two robots. We randomly assigned participants to play the game with either of the robots. We counterbalanced the order in which (1) the robot chose items and (2) the gaze manipulation appeared.

3.3 Hypotheses

Drawing from existing theory on nonverbal communication, we developed two main hypotheses on how the gaze cue would affect people's task performance and how the interpretation of the cue would differ between interactions with Robovie and 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 do so.

Hypothesis 2. – 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. Therefore, we expect the gaze cue to significantly affect task performance with Geminoid and not with Robovie.

3.4 Experiment Procedure

We first provided participants with a brief description of the purpose and procedure of the experiment. We 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. We deliberately concealed the primary purpose of our experiment—participants were not given any information on the robots' behavior. After the introduction, participants reviewed and signed a consent form and filled in a pre-experiment questionnaire on their affective state. We then provided them with more detail on the experimental task; the experimenter asked them to read a written description of the game and provided further detail on the task verbally. We then took participants into the experiment room to play the game with either Robovie or Geminoid. After playing a practice round, participants played eight rounds of the game. At the end of the game, we took were them out of the experiment room and asked them to fill in a post-experiment questionnaire that measured their affective state, personality, experience with and perceptions of the robot, perceptions of the task, and demographic information. Finally, the experimenter interviewed all participants regarding their experience.

The game task and the total experiment procedure took approximately 15 minutes and 45 minutes respectively. We conducted the experiment 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 ¥ (roughly $14 or €9) for their participation including their travel expenses.

3.5 Measurement

Our experimental design involved two manipulated independent variables, (1) whether or not the robot produced the gaze cue (manipulated as within-participants), and (2) whether they played the game with Robovie or Geminoid (manipulated as between-participants). The dependent variables involved *objective* and *subjective* measurements.

Objective – We measured participants' task performance through capturing the time it took participants and the number of questions they asked to identify the robot's picks. All sessions were videotaped to support the analysis of the objective measures.

Subjective – Subjective measures evaluated participants' affective state using the PANAS scale [55], perceptions of the robot's physical, social, and intellectual characteristics using a scale developed for evaluating humanlike agents [44] and attributions of mind and intentionality to the robot, perceptions of the task (e.g. how much they enjoyed and attended to the task), personality using scales of intellectual competence, creativity, distrust, and empathy [32], and demographic information. We measured participants' affective state before and after participants interacted with the robot and all other measurements after the experiment..We used seven-point Likert scales in all questionnaire items. We did the manipulation check 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 that they used in identifying the robots' picks. We 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.

3.6 Participation

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). Subjects were chosen to represent a variety of university majors. Of all the subjects, 11 studied engineering, 9 studied social sciences & humanities, 3 studied engineering, 2 studied natural sciences, and one participant did not report university major. The computer use among participants was very high (M=6.50, SD=0.65) on a scale from one to seven. Their familiarity with robots was relatively low (M=2.81, SD=1.55), so was their video gaming experience (M=30, 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 one ferret). Figure 4 shows participants playing the game with Robovie and Geminoid.

Figure 4. Participants in the experiment interacting with Robovie (top) and with Geminoid (bottom).

4. RESULTS

We analyzed objective measures using a mixed-effects analysis of variance (ANOVA). We included in the model participant and trial IDs as random effects and measured and manipulated independent variables (participant gender, pet ownership, and the robot with which participants interacted) as fixed effects. We analyzed subjective measures using a fixed-effects analysis of variance. We did the manipulation check using a contingency analysis. We also conducted correlation analyses to understand how subjective and objective measures correlated with each other.

Objective Measures – We used two main objective measures: the number of questions participants asked and the time it took them to identify the robot's picks. The number of questions provides us with an abstract measure of performance that indirectly quantifies the cognitive activity required to complete the task. However, the time required to identify the item might be a more accurate measure of participants' performance, because our observations during the two pretests that we conducted showed that even when participants had some idea which item the robot had in mind (inferred from their nonverbal behavior), they asked further questions to eliminate less possible alternatives, but did so without spending much time for cognitive processing. The task performance data included 208 trials. Two of these trials were excluded due to operator error. We also carefully studied the distributions and excluded 2 and 13 outliers that lied above 1.5 interquartile ranges (1.5·IQR) beyond the third quartile (Q_3) in the number of questions participants asked and the time it took participants to identify the item respectively. The resulting performance data included 200 and 193 trials for the former and latter performance measures respectively.

Our 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. Analyses of variance of both performance measures supported this hypothesis. Participants asked significantly fewer questions ($F[1,164]=4.30$, $p=0.04$) and took significantly less time ($F[1,150]=5.49$, $p=0.02$) to identify the robots' picks when the robots produced the gaze cue than when they did not do so (Figures 5.a and 5.b).

Our second hypothesis predicted that the gaze cue would affect participant performance with Geminoid but not with Robovie. Our analysis of the second performance measure provided support for this hypothesis. Participants identified the item significantly faster in the presence of the gaze cue when they played the game with Geminoid ($F[1,149]=3.93$, $p=0.05$), but their performance was not significantly affected by the gaze cue when they played the game with Robovie ($F[1,151]=1.75$, p=ns), as shown in Figure 5.c. On the other hand, 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 ($\chi^2(1,26)=7.54$, $p<.01$), as shown in Figure 5.d. Furthermore, our analysis showed that those who reported identifying the gaze cue did not differ in performance from those who did not report identifying the gaze cue ($F(1,22)=1.68$, p=ns). These findings are further supported by our 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.

Our analysis also showed that participants generally identified the item significantly faster with Robovie than with Geminoid ($F[1,23]=8.11$, $p<0.01$) as shown in Figure 5.c. This effect was present both when the robots produced the gaze cue ($F[1,46]=4.36$, $p=0.04$) and when they did not ($F[1,46]=7.06$, $p=0.01$). We discuss alternative explanations of this result in the Discussion section.

Our analysis found no effect of gender on how the gaze cue affected participants' performance but found a significant interaction between pet ownership and how the gaze cue affected the it took participants to identify the robots' picks ($F[1,174]=5.53$, $p=0.02$). Those who owned pets identified the robots' picks using significantly fewer questions ($F[1,173]=9.46$, $p<0.01$) and in a

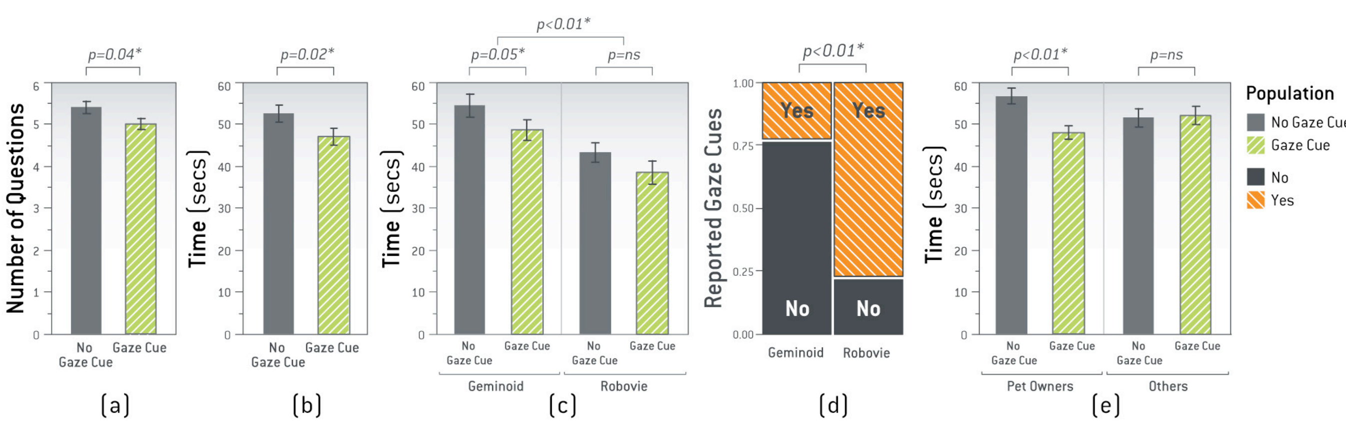

Figure 5. Results on objective measures: (a) Number of questions that participants asked to identify the item with and without gaze cue, (b) the time it took participants to identify the item with and without gaze cue, (c) the time it took participants to identify the item with and without the gaze cue for the two robots, (d) whether or not participants reported identifying the gaze cue for the two robots, and (e) the time it took pet owners and others to identify the item with and without gaze cues. Lower ratings represent better task performance. (*) denotes statistically significant probabilities.

significantly shorter time (F[1,158]=8.84, p<0.01) when the robots produced the gaze cue than when they did not do so (Figure 5.e). Those who did not own pets showed no differences in the number of questions that they asked (F[1,172]=0.06, p=ns) and the time it took them (F[1,160]=0.21, p=ns) to identify the robots' picks with the presence of the gaze cue.

Subjective Measures – We started our analysis of subjective measures with a factor analysis of 30 questionnaire items that we used to evaluate social and intellectual characteristics of the robots. The analysis produced eight factors from which we created two reliable measures: 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 more socially desirable than they rated Geminoid (F[1,20]=12.49, p<0.01). We also found a significant interaction effect between participant gender and robot (F[1,20]=7.79, p=0.01). Women rated Robovie as significantly more socially desirable than they rated Geminoid (F[1,20]=14.95, p<0.01), while no differences were found in men's ratings of the social desirability of the two robots (F[1,20]=0.29, p=ns). The analysis also produced a marginal interaction effect between pet ownership and robot on participants' ratings of the social desirability of the robots (F[1,20]=3.21, p=0.09). Those who did not own pets rated Robovie as more socially desirable than they rated Geminoid (F[1,20]=15.30, p<0.01) while pet owners did not differ in their evaluations of the two robots (F[1,20]=1.46, p=ns).

No differences in participants' ratings of the two robots' intelligence and their attributions of mind to the robots were observed (F[1,20]=1.91, p=ns). This result is consistent with the qualitative data obtained through interviews as participants mainly associated intelligence with the robots' behavior—that the robots could answer all of their questions in the game—and not with their physical design.

Some of the factors in our factor analysis were loaded on single items. Therefore, we also analyzed single items using analyses of variance. No differences were observed in how much participants liked the robot (F[1,20]=0.52, p=ns) or how much they thought that the robot liked them (F[1,20]=1.24, p=ns). However, both measures were affected by whether participants owned pets. We found a marginal interaction between pet ownership and which robot participants interacted with in how much they liked the robots (F[1,20]=3.16, p=0.09) and a significant interaction between the same independent variables in how much they thought that the robot liked them (F[1,20]=6.68, p=0.02). While pet owners did not differ in their ratings of how much they liked the two robots (F[1,20]=0.51, p=ns) and of how much they thought that the robot liked them (F[1,20]=0.99, p=ns), those who did not own pets liked Robovie marginally more (F[1,20]=3.33, p=0.08) than Geminoid and thought that Robovie liked them significantly more than they thought Geminoid did so (F[1,20]=7.31, p=0.01).

We also conducted multivariate analyses of our objective and subjective data to understand how technology use and personality measures correlated with task performance. We found no significant correlations between performance measures and technology use. We found a significant negative correlation between the number of questions participants asked to identify the robots' picks and participants' trust (r=-0.18, p<0.01) and between the time it took participants to identify the item and participants' comprehension (r=-0.16, p= 0.03), initiative (r=-0.17, p=0.02), and quickness (r=-0.15, p=0.05).

5. DISCUSSION

The results supported our first hypothesis. Participants performed better in two performance measures when the robots "leaked what they had in mind" by means of gaze than when they did not do so, from which we infer that they read the leakage cue, and attributed mental states and intentionality to the robots, and used this information in their task. Our second hypothesis was also 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. We also found that participants were more likely to report identifying the gaze cue with Geminoid than with Robovie and that those who reported identifying the gaze cue did not differ in their performance from those who reported identifying the gaze cue, supporting the argument that people automatically and unconsciously read and respond to leakage cues.

We also found strong effects of pet ownership on all objective measures. Gaze cues affected only pet owners' performance in the game and not others, suggesting perhaps that people who own pets might become—through their interaction with their pets—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 intentions [39]. Research on embodied virtual agents has also shown that dog owners differed from others in how they evaluated agents with zoomorphic features [44].

Design and Research Implications – While the work presented here is a first step towards understanding how robots might use seemingly unintentional cues to communicate intentions, it has a number research and design implications for human-robot interaction. Nonverbal leakages—and, more broadly, seemingly unintentional behavior—might provide the design of humanlike robots with opportunities to create rich, humanlike behavior. For instance, a shaking limb might communicate nervousness more expressively than explicit facial or verbal expressions. This work also informs research in shared attention and theory of mind in human-robot interaction. Our study showed that even two 400-millisecond glances could lead to establishing shared attention, attribution of intentionality, and task performance effects. Furthermore, this work extends our understanding of how people interpret and respond to human communicative cues when they are used by robots.

Limitations – The within-participants design of our experiment limited our ability to measure the effect of the gaze cue on subjective evaluations of the robot. While we deliberately chose this design to account for some of the variability in participants' task performance that individual differences might cause, we acknowledge the importance of gaining a better understanding of how leakage cues might affect subjective attributions of intentionality, purposefulness, and states of mind. Therefore, we plan to run a follow-up study using the same task in a between-participants design.

Our 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 interacting with Geminoid was cognitively and perceptually more demanding than interacting with Robovie was. Data from our semi-structured interviews provides some support for this explanation. Participants consistently reported being surprised by how humanlike the robot looked. They also reported feeling nervous, lose focus, and get distracted from their task. Two participants reported that they could not relate to the robot because it looked older than them, suggesting an alternative explanation for why participants performed more poorly with Geminoid than with Robovie; they might have used polite language in talking to Geminoid, following Japanese conversational conventions, which would take more time for cognitive processing and language construction. This would adversely affect performance as we measured it in our study.A content analysis of the transcripts from the video data would verify this explanation. We plan to further analyze our data int he future. Participants also reported their nervousness diminished over time, suggesting that allowing participants to interact with Geminoid in an immersive and non-intimidating task before they performed in the experiment might have alleviated some of the effects caused by the design of the robot. We plan to employ this approach in our future work with Geminoid.

In building our hypotheses and designing the experiment, we assumed that there would be no significant differences in the accuracy of participants' perception of the two robots' gaze direction (whether participants can identify the item toward which the robot is looking). While we carefully designed the two robots' gaze behaviors to be identical, paying particular attention to precision, there might be inherent differences in how the human communicative system responds to the gaze cues produced by the two robots. To validate this possibility, we conducted a follow-up experiment in which we compared how accurately participants interpreted the two robots' gaze direction. We also added a human confederate in the comparison to gain a better understanding of whether the accuracy of people's perception of the gaze directions of either robot differed significantly from that of human gaze direction. In a within-participants-design experiment, we asked 12 naïve participants (5 males and 7 females with an average age of 20.1, ranging between 18 and 22) to rate the gaze target of (1) Robovie, (2) Geminoid, and (3) our human confederate as they glanced towards a randomly selected item from among the 14 objects used in our experimental task. Each participant performed the task in 12 trials for each condition in a counterbalanced order. Participants rated Robovie's, Geminoid's, and the human confederate's gaze directions with an average accuracy of 31.94% (SD=17.71%), 39.58% (SD=12.37%), and 37.50% (SD=15.28%) respectively with a baseline accuracy of 7.14% (for random guess). These results are also illustrated in Figure 6. We conducted a random-effects analysis of variance (ANOVA) and found that the overall model was not significant ($F[2,20]=0.51$, p=ns). Similarly, pairwise comparisons produced no significant differences among the accuracy ratings of the three gaze sources.

In summary, both in the presence and absence of the gaze cue, overall, participants performed better with Robovie than with Geminoid. We argue 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. We 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., determine accurately the directions of gaze) and correctly interpret it (i.e., attribute intentionality and use this information to improve their performance in their task). We plan to further analyze our data and conduct follow-up studies to concretize these explanations.

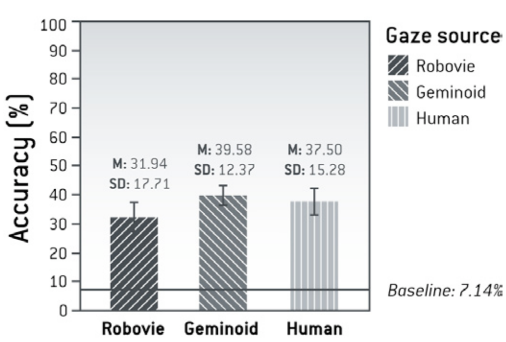

Figure 6. Comparison of how accurately participants perceived the gaze direction of Robovie, Geminoid, and a human confederate. No statistical differences were observed in accuracy across gaze sources.

CONCLUSIONS

Human communication involves a number of nonverbal cues that are produced unintentionally and communicate a wealth of information on the mental state and intentions of individuals. Leakage cues are a particular set of such cues that "leak" information on mental states, emotions, and intentions through the nonverbal channel. In this paper, we 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. In a controlled laboratory study, we showed that participants performed better when the robots *leaked* information through cues as minimal as two 400-millisecond glances, from which we infer that they read these cues, interpreted these cues as related to their task, and used this information to improve their performance. We compared two robots with different levels of humanlikeness, Geminoid—a near-human android—and Robovie—a humanoid robot with abstract, stylized humanlike features—in how the production of leakage cues affected participants' task performance in a guessing game. We found that the presence of the cues led to significant improvements in their task performance only with Geminoid, which might suggest that more humanlike faces are more appropriate to communicate intentions and mental states through leakage cues. We also found that fewer participants reported identifying the leakage cue with Geminoid than with Robovie, suggesting a more automatic and unconscious response to the cues produced by Geminoid than those by Robovie. Furthermore, whether or not they reported identifying the gaze cue did not affect their performance, further supporting the argument that people automatically and unconsciously read and respond to leakage cues. We found that the leakage cue affected the performance of only pet owners and not others, which might suggest that pet owners become—through their interaction with their pets—more sensitive to nonverbal behavior

While this study is a first step in understanding the role that seemingly unintentional cues might play in human-robot interaction, it provides evidence that these cues can be used by robots to communicate mental states and intentions. Further work is required to extend these results into design guidelines and to better understand how robot characteristics such as humanlikeness shape people's judgments of nonverbal cues in robots.

ACKNOWLEDGEMENTS

This work supported in part by the JSPS Grant-in-Aid for Scientific Research (S), KAKENHI (20220002). The first author was supported in part by NSF grant IIS-0624275. We would like to thank Sara Kiesler and Jodi Forlizzi for their feedback on experimental design, Susan Fussell and Lindsay Jacobs for thier valuable comments on our paper, and Kazuhiko Shinozawa for his support with the implementation of the experiment.

REFERENCES

[1] Ambady, N. and Rosenthal, R. Thin Slides of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. *Psychological Bulletin*, 111 (2), 256-274, 1992.

[2] Argyle, M. et al. The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European J. Social Psychology*, 1 (3), 385-402, 1971.

[3] Bailenson J. N. et al. Detecting digital chameleons. *Computer in Human Behavior*, 24, 66-87, 2008

[4] Baron-Cohen, S. *Mindblindness: An Essay on Autism and Theory of Mind.* MIT Press, 1995.

[5] Bartneck, C. et al. In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions. *In Proc. D&E'04,* 2004.

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

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

[8] Breazeal, C. and Scassellati, B. How to build robots that make friends and influence people. *In Proc. IROS*, 1999.

[9] Breazeal, C. Emotive qualities in robot speech. *In Proc. IROS'01*, 2001.

[10] Breazeal, C. *Designing Sociable Robots*, MIT Press, 2002.

[11] Bruce, A. et al. The role of expressiveness and attention in human-robot interaction. *In Proc. ICRA'02*, 4138-4142, 2002.

[12] Buller, D. B. et al. Testing interpersonal deception theory: The language of interpersonal deception. *Comm. Theory*, 6, 268-289, 1996.

[13] Calder, A. J. et al. Reading the mind from eye gaze. *Neuropsychologia*, 40, 1129-1138, 2002.

[14] Castelli, F. et al. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125, 1839-1849, 2002.

[15] Castiello, U. Understanding other people's actions: Intention and attention. *J. Exp. Psych.: Human Perc. & Perf.,* 29, 416-430, 2003.

[16] Chartrand, T. L. and Bargh, J. A. The chameleon effect: the perception-behavior link and social interaction. *J. Pers. & Soc. Psych.*, 76, 893-910, 1999.

[17] Chawla, P. and Krauss, R. M. Gesture and speech in spontaneous and rehearsed narratives. *J. Exp. Soc. Psych.,* 30 (6), 580-601, 1994.

[18] DePaulo, B. M. et al. Cues to Deception. *Psych. Bulletin.* 129 (1), 74-118, 2003.

[19] Ekman, P. and Friesen, W.V. Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88-106, 1969.

[20] Ekman, P. et al. The Duchenne smile: Emotional expression and brain physiology II. *J. Pers. & Soc. Psych.*, 58, 342-353, 1990.

[21] Emery, N. J. The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosc. & Biobehav. Rev.*, 24 (6), 581-604, 2000.

[22] Feldman, R. S. et al. Nonverbal Cues as Indicators of Verbal Dissembling*. American Ed. Res. J.,* 15 (2), 217-231, 1978.

[23] Freire, A. et al. Are Eyes Windows to a Deceiver's Soul? Children's Use of Another's Eye Gaze Cues in a Deceptive Situation. *Dev. Psych.*, 40 (6), 1093-1104, 2004.

[24] Frischen, A. et al. Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. *Psych. Bulletin,* 133 (4), 694-724, 2007.

[25] Hemsley, G. D. and Doob, A. N. The effect of looking behavior on perceptions of a communicator's credibility. J. *Appl. Soc. Psych.,* 8, 136-144, 1978.

[26] Hinds, P. J. et al. Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *HCI J.*, 19 (1), 151-181, 2004.

[27] Imai, M. et al. Robot mediated round table: Analysis of the effect of robot's gaze. *In Proc. ROMAN'02*, 411-416, 2002.

[28] Imai, M. et al. Physical relation and expression: joint attention for human-robot interaction. *IEEE Trans. Ind. Elec.,* 50 (4), 636-643, 2003.

[29] Gockley, R. et al. Interactions with a moody robot. *In Proc. HRI'06,* 2006.

[30] Goetz, J. et al. Matching robot appearance and behavior to tasks to improve human–robot cooperation*. In Proc. ROMAN'03*, 2003.

[31] Goffman, E. *Relations in public. New* York: Harper & Row, 1971.

[32] Goldberg, L. R. et al. The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality, 40, 84-96, 2006.

[33] Kanda, T. et al. Body Movement Analysis of Human-Robot Interaction. *In Proc. IJCAI'03*, 177-182, 2003.

[34] Kiesler, S. and Goetz, J. Mental models of robotic assistants. *In Proc. CHI'02*, 576-577, 2002.

[35] Kleinke, C. L. Gaze and eye contact: A research review. *Psychological Bulletin*, 10, 78-100, 1986.

[36] Krauss, R. M. et al. Verbal, vocal and visible factors in judgments of another's affect. *J. Pers. & Soc. Psych.*, 40, 312-320, 1981.

[37] Kraut, R. E. and Poe, D. Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *J. Pers. & Soc. Psych.*, 39, 784-798, 1980.

[38] Leslie, A. M. Pretense and representation: The origins of "theory of mind". *Psych. Review*, 94, 412-426, 1987.

[39] Miklósi, A. et al. Intentional behavior in dog-human communication: An experimental analysis of "showing" behavior in the dog. *Animal Cognition*, 3 (3), 159-166, 2000.

[40] Miwa, H. et al. Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII. *In Proc. IROS'04*, 2004.

[41] Mutlu, B. et al. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *In Proc. Humanoids*, 2006.

[42] Nowak, K. Defining and differentiating copresence, social presence and presence as transportation. *In Proc. Presence*, 2001.

[43] Nowak, K. L. and Biocca, F. 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, 2003.

[44] Parise, S. et al. Cooperating with life-like interface agents. *IBM WRC Technical Report,* 1998.

[45] Perrett, D. I. and Emery, N. J. Understanding the intentions of others from visual signals: neurophysiological evidence. *Cahiers de Psychologie Cognitive,* 13, 683-694, 1994.

[46] Pierno, A. C. et al. When gaze turns into grasp. *J. Cog. Neuroscience,* 18, 2130-2137, 2006

[47] Scheeff, M. et al. Experiences with Sparky: A social robot. *In Proc. Workshop on Interactive Robot Entertainment*, 2000.

[48] Scherer, K. R. et al. Minimal cues in the vocal communication of affect: Judging emotion from content-masked speech. *J. Psycholinguistic Research*, 1, 269-285, 1972.

[49] Scherer, K. R. et al. The voice of confidence: Paralinguistic cues and audience evaluation. *J. Res. in Pers.,* 7, 31-44, 1973.

[50] Sidner, C. L. et al. Where to look: a study of human-robot engagement. *In Proc. IUI'04,* 2004.

[51] Sugiyama, O. et al. Three-layered draw-attention model for humanoid robots with gestures and verbal cues. *In Proc. IROS'05*, 2005.

[52] Surakka, V. and Hietanen, J. K. Facial and emotional reactions to Duchenne and non-Duchenne smiles. *Intl. J. Psychophysiology*, 29, 23-33, 1998.

[53] Tojo, T. et al. A conversational robot utilizing facial and body expressions. *In Proc. Systems, Man, and Cybernetics'00*, 2000.

[54] Trafton, J. G. et al. Integrating vision and audition within a cognitive architecture to track conversations. *In Proc. HRI'08*, 2008.

[55] Watson D. et al. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers. Soc. Psychol*., 54 (6), 1063-1070, 1988.

[56] Waxer, P. H. Nonverbal cues for anxiety: an examination of emotional leakage. *J. Abnormal Psych.*, 86 (3), 306-314, 1977.

[57] Williams, L. M. et al. In Search of the "Duchenne Smile": Evidence from Eye Movements. *J. Psychophysiology,* 15 (2), 122-127, 2001.

[58] Zuckerman, M. et al. Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, 14, 1-59. New York: Academic Press, 1981.