Speech and Gaze Conflicts in Collaborative Human-Robot Interactions

Henny Admoni, Christopher Datsikas, and Brian Scassellati
(henny@cs.yale.edu, christopher.datsikas@yale.edu, scaz@cs.yale.edu)
Department of Computer Science, Yale University
New Haven, Connecticut 06520 USA

Abstract

Gaze and speech are both important modes of communication for human-robot interactions. However, few studies to date have explored the effects of conflict in a robot's multi-modal communication. In this paper, we investigate how such speech-gaze conflicts affect performance on a cooperative referential task. Participants play a selection game with a robot, in which the robot instructs them to select one object from among a group of available objects. We vary whether the robot's gaze is congruent with its speech, incongruent with its speech, or absent, and we measure participants' response times to the robot's instructions. Results indicate that congruent speech facilitates performance but that incongruent speech does not hinder performance. We repeat the study with a human actor instead of a robot to investigate whether human gaze has the same effect, and find the same results: in this type of activity, congruent gaze helps performance while incongruent gaze does not hurt it. We conclude that robot gaze may be a worthwhile investment in such situations, even when gaze behaviors may be unreliable.

Keywords: human-robot interaction; eye gaze; non-verbal communication

Introduction

In typical human interactions, eye gaze supports and augments spoken communication (Kleinke, 1986). People gaze almost exclusively at task-relevant information (Hayhoe & Ballard, 2005), and gaze is used to disambiguate statements about objects in the environment (Hanna & Brennan, 2007). Similar mechanisms are also at play in human-robot interactions: task-relevant robot eye gaze can be used to improve the efficiency of collaborative action (Boucher et al., 2012).

For example, imagine a human and robot collaboratively constructing a birdhouse. The robot can use its eye gaze to clarify an ambiguous speech reference, saying "Please pass the green block" while looking at a particular green building block to distinguish it from among other green blocks. This multi-modal communication makes the interaction more efficient by using multiple channels to convey information, requiring less investment in costly mechanisms like generating sufficiently descriptive speech, and improving the naturalness of the interaction (Huang & Thomaz, 2011).

But robots are not perfect, and sometimes speech and gaze cues will conflict. Sensor errors, hardware malfunctions, and software bugs can cause mismatches between a robot's gaze and speech. In such cases, a human partner receives incorrect or contradictory information from the robot. The human might misinterpret the robot's speech or, at best, must hesitate to decide what the robot means, decreasing the collaboration's efficiency and increasing the human's cognitive load.

While a growing body of evidence shows that people can interpret robot gaze and speech, only a few studies to date have investigated the effects of speech-gaze conflicts. In this paper, we investigate how speech-gaze conflicts are handled by human partners in collaborative, embodied human-robot interactions. We focus on object selection tasks in which a robot provides instructions to a human, because these scenarios are central to collaborative action, and because communication misinterpretation in such scenarios can be costly.

We compare *congruent* gaze—in which the robot looks at the object it references in speech—and *incongruent* gaze—in which the robot looks at a different object—to a control condition in which the robot does not exhibit gaze cues. To quantitatively measure the effect of speech-gaze conflicts, we record the time between when the robot begins its instructions and when participants select an object. Response time serves as an approximation of task efficiency; faster responses mean less overall time taken for the task.

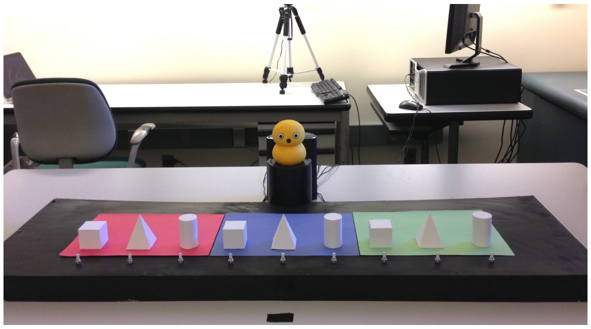
As a final manipulation, we also include a human agent condition, in which the robot is replaced by a person who performs the robot's role in the experiment. The human agent condition attempts to discover whether robot gaze is any more or less influential on human behavior than human gaze.

The results of this study provide evidence of the effectiveness of gaze in collaborative human-robot interactions. As described below, we find that congruent gaze facilitates performance in both robot and human conditions. Interestingly, we also find that incongruent gaze does not hinder performance in either the robot or the human conditions. In other words, in this task, people are able to recover quickly enough from speech-gaze conflicts that their performance is statistically no different than not having gaze at all. These results suggest that adding referential gaze may be a low-risk way to improve human performance in similar environments, even when the gaze system is unreliable.

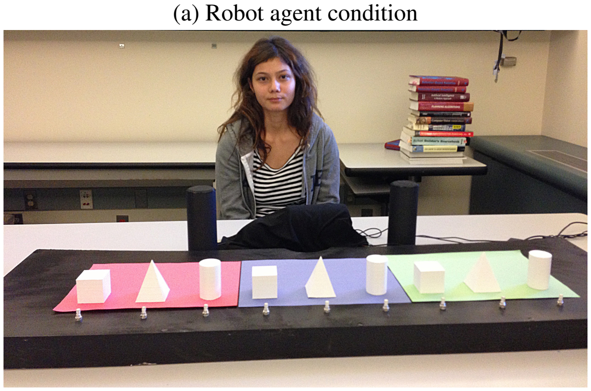
Related Work

Directional eye gaze seems to be a special stimulus, evoking reflexive attention shifts that are robust to top-down modulation (Friesen, Ristic, & Kingstone, 2004). Functional MRI studies reveal a significant overlap in the brain areas that process theory of mind and those that process eye gaze (Calder et al., 2002). In fact, observing someone signaling the presence of an object with referential gaze elicits the same neural response as observing someone physically reaching to grasp that object (Pierno et al., 2006), indicating that people use gaze as a powerful indicator of others' future behavior.

Where we look is closely coupled with what we say in human-human interactions. Objects or figures in the environment are typically fixated one second or less before they



(a) Robot agent condition



(b) Human agent condition

Figure 1: Participant view of the experiment. MyKeepon or a human actor provided verbal and gaze cues about which shape to select.

are named in conversation (Griffin & Bock, 2000; Yu, Schermerhorn, & Scheutz, 2012). When referencing objects, people use eye gaze as a strong and flexible cue for eliminating ambiguity (Hanna & Brennan, 2007). When access to a partner's eye gaze is restricted, for instance because the partner is wearing sunglasses, people are slower at responding to their partner's referential communication (Boucher et al., 2012).

As in human-human interactions, eye gaze is an important part of human-robot interactions. Robot eye gaze can influence whether people join a conversation or feel excluded from it (Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009), can influence people to favor certain objects over others (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009), and can facilitate cooperative behaviors like object handoffs between humans and robots (Strabala et al., 2013). Exhibiting joint attention, a type of social gaze, increases ratings of a robot's competency and naturalness (Huang & Mutlu, 2012).

More specifically, studies of human-robot interaction have shown that robot gaze can be used to clarify speech. If a robot gazes toward an object while naming it, people select the object more quickly than if the robot names the object without looking at it (Boucher et al., 2012; Huang & Mutlu, 2012). With both robots (Huang & Mutlu, 2012; Mutlu, Forlizzi, & Hodgins, 2006) and virtual agents (Andrist, Pejsa, Mutlu, & Gleicher, 2012), gazing at task-relevant objects during teaching—for instance, looking at a map while describing political boundaries—increases peoples' retention of information. In most of the literature about referential gaze in HRI, however, robot gaze is *congruent* with speech.

Some researchers have investigated the effects of speech and gaze conflicts in HRI. In a video-based study (Staudte & Crocker, 2011), participants evaluated the correctness of a robot's statements about objects in front of it (for instance, "the cylinder is bigger than the pyramid that is pink"). When the robot's gaze was congruent with its speech, response times were shorter than a no-gaze control; when gaze was incongruent, response times were longer than the control. This suggests that people relied on gaze to facilitate sentence processing, and that incongruent gaze hinders comprehension.

Unlike our experiment, however, Staudte and Crocker's task involved sentence evaluation rather than object selection, which requires a different cognitive skill set. Furthermore, their study was conducted with video stimuli instead of embodied robots. While virtual robots increase the ease of use and replicability of stimuli, they may not have as strong an influence on human behavior as physically embodied robots (Bainbridge, Hart, Kim, & Scassellati, 2011).

In contrast, research using an object selection task and an embodied robot finds no difference in response times between no gaze and incongruent gaze conditions, though results support the benefit of congruent gaze (Huang & Mutlu, 2012). However, this study used a between-subjects design in which the robot exhibited only one type of gaze (congruent or incongruent) to each participant. Participants could acclimate to the robot's gaze strategy, which does not address situations where gaze is usually helpful but occasionally incorrect.

The current work is inspired by these studies, and builds upon them by investigating conditions in which speech and gaze are incongruent rather than only congruent (Boucher et al., 2012), using a physically embodied robot rather than a video (Staudte & Crocker, 2011), and introducing uncertainty about the robot's reliability to avoid habituation to one particular condition (Huang & Mutlu, 2012).

Experiment 1

This experiment is designed to investigate whether gaze conflicts hinder task performance in collaborative human-robot interactions. Participants engaged in an object selection task with a robot. On each trial, the robot provided spoken instructions of the form "Please pick the [shape] in the [color] zone" where shape and color referred to objects in front of the participant (Figure 1). Each of the nine objects was referenced nine times during the interaction, for a total of 81 trials.

On each trial, the robot also provided a gaze cue, which was either congruent with the speech (i.e., looking at the same object), incongruent with the speech (i.e., looking at a different object), or absent (no movement). The robot started each trial in a neutral position, with gaze directed straight forward and approximately 30 cm below the participant's eyes. To initiate a gaze cue, the robot first attempted to establish joint attention by looking up at the participant's face (mutual gaze), and then engaged in object reference by looking down toward

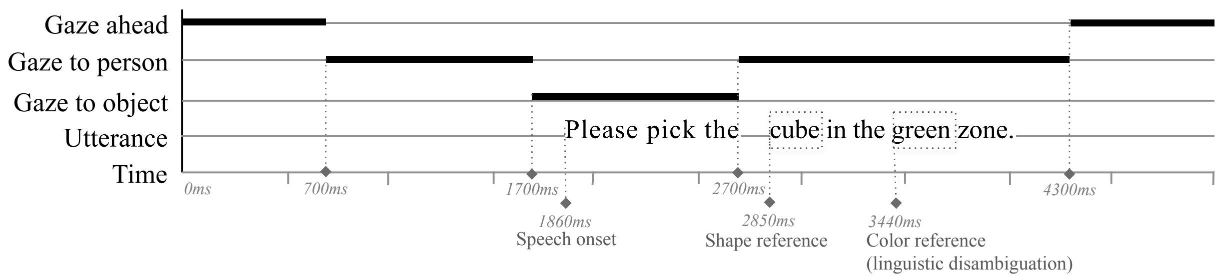


Figure 2: A visual representation of the speech and gaze during a typical trial. This figure shows a congruent trial: the agent both verbally and physically indicates the green cube. In an incongruent trial, the spoken word "green" is replaced with one of the other two zone colors. In a no-gaze trial, the agent gazes straight ahead.

the selected object, then returned to look at the participant's face before returning to the neutral position (Figure 2). The robot did not use sensors to confirm that mutual eye gaze was successful; instead, the experiment was pre-scripted and ran autonomously. In no gaze trials, the robot did not move at all and continued looking ahead in neutral position.

In human-human communication, eye gaze moves toward an object prior to a verbal reference and away from the object just as it is named (Yu et al., 2012). We carefully aligned the verbal and gaze cues to mimic natural behavior (Figure 2).

We measured how much time participants took to select a shape. By comparing response times in the congruent, incongruent, and no gaze conditions, we are able to determine whether gaze has any facilitation or hindrance effect on the speed with which people respond to the robot's instruction.

Apparatus

The experiment apparatus is a black box measuring 120cm by 40cm by 6cm (Figure 1). The robot was placed on the table across the box from the participant, approximately 80 cm away. Three zones are marked by colored paper on top of the box: red, blue, and green. Each zone contains an identical set of white blocks in simple shapes—a cube, a pyramid, and a cylinder—arranged side-by-side in a single row on top of the zone. A momentary pushbutton switch in front of each object is used to select that object, and the precise timings of button presses are recorded on a nearby computer.

We used a relatively inexpensive, easily modifiable, and commercially available robot called MyKeepon (Figure 1). This 32cm tall, snowman shaped, interactive robot toy has a rubber yellow skin and four degrees of freedom: rotation around the base, left/right lean, front/back lean, and up/down bob. MyKeepon is a consumer-grade version of a research robot called Keepon Pro, which was designed to be a socially evocative but simple robotic agent (Kozima, Michalowski, & Nakagawa, 2009). The robot's minimalist design and salient eyes make it a useful platform for HRI studies about eye gaze.

We modified a MyKeepon to make it programmable for this experiment. The MyKeepon internal microprocessors were connected to an Arduino Uno, an open-source electronic prototyping platform. Using the I2C bus on the MyKeepon microprocessor and open-source software (Michalowski, Machulis, & Gasson, 2013), the Arduino sends motor commands and retrieves information such as encoder positions from the MyKeepon hardware. This allowed for easy control of the MyKeepon robot platform.

Procedure

Twenty two people participated in this experiment (10 females). Their ages ranged from 18 to 34, with a mean age of 22, and most were Yale undergraduate students. Participants were compensated $8.

Participants were told that they would play an object selection game to help evaluate a new robot platform called MyKeepon. They were shown the nine shapes and told that in each round of the game, MyKeepon would provide instructions on which shape to choose. Participants were informed that they should select the shape as quickly and accurately as possible. They were also told to return their finger to a marked start position on the table between trials. This instruction was given to eliminate any "hovering" over the buttons so that response times are consistent across trials.

In each trial, a computer-generated voice provides a verbal cue, which is a sentence that first indicates the type of object and the zone the object is in, for example, "Please pick the cube in the green zone" (Figure 2). The sentence is constructed so that the specific object referred to by the sentence remains ambiguous until the color of the zone is stated near the end of the sentence. Until this point of linguistic disambiguation, there are three potential matches for the sentence (the named shape in the red, blue, and green zones), so participants cannot select an object with more than 33% reliability.

Simultaneous with the verbal cue, the robot also provides a gaze cue by orienting toward one of the shapes. On congruent trials, the robot turns toward the shape named by the verbal cue. On incongruent trials, the robot turns toward a different shape at least three spots away from the correct shape. This restriction ensures that there is no confusion about whether the gaze was directed toward the correct shape. On no gaze trials, the robot remains looking straight ahead.

Participants first practiced two congruent gaze trials under experimenter supervision to familiarize themselves with the task; these practice trials were not recorded, and participants were not told that the robot's gaze would vary in other trials. After the practice, each participant experienced two sections of the experiment with no breaks between them.

In the first section, called the *blocked* section, participants saw each trial type blocked together: first nine no gaze trials, then nine congruent trials, then nine incongruent trials, with no demarcation between the blocks. The purpose of the blocked section is to establish a baseline measure of reaction time (in the no gaze block) and to observe how performance changes as participants become familiar with the robot's gaze.

The second section of *randomized* trials followed the blocked section immediately. During the randomized section, each participant saw a unique random ordering of all 54 combinations of shape, color zone, and gaze type. The purpose of this section is to measure the effects of gaze cues when participants did not know whether the cue would help or not.

After both sections, participants were given a survey with demographic questions and one free-response question: "Did you notice anything unusual about the robot's behavior?"

Results

Twenty-two participants each completed 27 blocked trials and 54 randomized trials for a total of 1782 data points. Four trials (0.2%) were discarded because no response was recorded within 12 seconds, either because the participant did not press a button or because the button press did not register. We also discarded the no gaze blocked section for one participant (nine trials, or 0.5% of all trials) due to self-reported noncompliance. Participants were highly compliant with the verbal cue, selecting a shape that was different from the robot's spoken instruction on only five trials (0.3%). Results are shown in Table 1 and Figure 3a.

A repeated measures ANOVA of response time by gaze type for all trials shows a significant main effect ($F(2,42) = 43.181, p < 0.001$). Post-hoc tests with a Bonferroni correction reveal that response times to congruent gaze were significantly shorter than response times to incongruent gaze (by 242ms, $p < 0.001$) and to no gaze (by 262ms, $p < 0.001$). There was no statistical difference between incongruent and no gaze trials.

There are several conclusions to be drawn from these results. First, people were highly accurate and highly consistent in following the robot's speech, complying with speech instructions on 99.7% of trials even though 33% of the trials included a conflicting gaze cue. The high rate of compliance with speech suggests that cases in which participants failed to follow the speech cue involved button press errors, though we did not explicitly ask participants to report errors.

When the robot's gaze indicated the same shape as the verbal cue, participants used gaze to guide their responses, as indicated by the significantly improved response times in the congruent gaze condition. Surprisingly, participants were not hindered by incorrect robot gaze: they responded no slower to incongruent trials—when gaze and speech did not match—than they did to no gaze trials, where there was no gaze cue. In other words, congruent gaze helped people respond to a robot's verbal instructions more quickly, but incongruent gaze did not make them respond more slowly than no gaze at all.

Agent	Gaze type	RT (ms)	SD (ms)	N
Robot	Congruent	4572	256	22
	Incongruent	4814	235	
	None	4834	222	
Human	Congruent	4427	309	9
	Incongruent	4568	289	
	None	4621	189	

Table 1: Response times (RTs) for all trials in Experiments 1 and 2. RTs are measured from start of trial, including time to speak sentence. When measured from linguistic disambiguation, RTs are similar to previous work (Boucher et al., 2012).

The response facilitation from robot gaze supports previous findings in HRI (such as Boucher et al., 2012; Huang & Mutlu, 2012; Staudte & Crocker, 2011). However, the lack of hindrance from incongruent gaze conflicts with previous findings (Staudte & Crocker, 2011).

To test whether this effect is due to the robot or to the task, we conduct a new experiment with a human in place of the robot. If the same procedure—now with human gaze—yields the same effect, we can conclude that the task, and not the agent, is responsible for the absence of hindrance.

Experiment 2

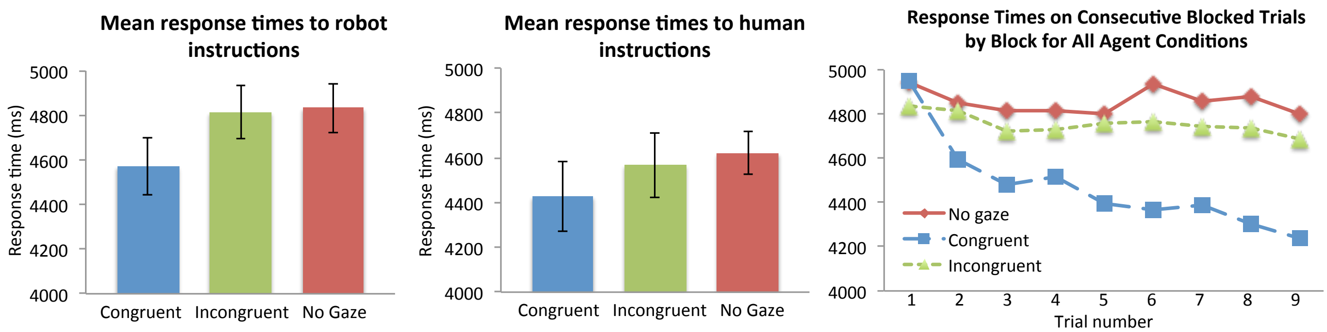
We replicated Experiment 1 with a small number of participants. The apparatus and procedure are identical to Experiment 1, except that the robot is replaced by a human actor (Figure 1b). For consistency, the verbal cue is still provided by the computer-generated voice from Experiment 1. We took care to make the human gaze as similar as possible to the robot gaze; therefore, the actor practiced looking at the object for the correct duration and shifting her gaze away from the referenced object just before it was named. On the post-task questionnaire, the free-response question was changed to: "Did you notice anything unusual during the experiment?"

Nine participants (2 females) took part in Experiment 2. Their ages ranged from 18 to 20 (mean of 19). They were all Yale undergraduates and they were compensated $8.

Results

Table 1 and Figure 3b show the results of Experiment 2. A repeated measures ANOVA to test the effect of gaze type on response times found a significant main effect ($F(2,16) = 7.892$, $p = 0.004$). Post-hoc tests with a Bonferroni correction reveal that response times to congruent gaze are shorter than response times to incongruent gaze (141 ms, $p = 0.018$) and no gaze (194 ms, $p = 0.033$). No significant difference was found between response times in incongruent and no gaze conditions. Participants made an erroneous selection (not following the speech cue) on 20 (2.7%) of the 729 trials.

To compare robot and human gaze, we conducted an ANOVA on response time with gaze type as a within-subjects factor and agent type as a between-subjects factor. The analysis reveals a significant effect of gaze ($F(2,86) = 6.564, p = 0.002$) but no effect of agent ($F(1,86) = 0.351, p = ns$) and



(a) Experiment 1, robot agent.

(b) Experiment 2, human agent.

(c) Blocked trials, both agents.

Figure 3: Response times to agent instructions. Figures (a) and (b) show mean response times across all trials for each experiment. Figure (c) shows the blocked trials section separated by trial number. In all figures, congruent gaze facilitates response times, while incongruent and no gaze conditions show no significant difference. Error bars show 1 standard error.

no interaction ($F(2,86) = 0.291, p = ns$). Post-hoc pairwise comparisons on the significant result show that congruent gaze led to shorter response times than incongruent gaze (191 ms, $p = 0.017$) and no gaze (228 ms, $p = 0.003$) for all participants regardless of agent condition. No significant difference was found between incongruent and no gaze conditions.

The blocked section of the experiment reveals how participants acclimated to a consistent gaze type. Because there is no significant difference between agent conditions, we can collapse the data across these conditions for this analysis. Figure 3c shows mean response times for each trial in the blocked section, averaged across participants in both agent conditions. Recall that participants saw nine no gaze trials, then nine congruent trials, and then nine incongruent trials. The no gaze block serves as a baseline for response times without gaze. As shown in Figure 3c, response times remained fairly stable during the no gaze block. Response times improved during the congruent block, indicated by the downward slope of the congruent block line. In contrast, there was no improvement of performance over the nine incongruent blocked trials. Participants performed slightly better on the incongruent block than on the no gaze block that preceded it, though this effect may be due to practice.

Although participants were never explicitly told to follow gaze, they rapidly adapted to using congruent gaze to improve their performance. The rate of improvement does not decrease by the ninth trial, suggesting that more congruent gaze trials might have led to continuing improvements.

General Discussion

For both robot and human agents, participant response times were faster when the agent's gaze cue was congruent with its verbal cue, compared to incongruent gaze and no gaze conditions. Because the gaze cue is delivered before the point of linguistic disambiguation, the fact that participants responded more quickly on congruent gaze trials indicates that they planned their motion according to the gaze cue before hearing the disambiguation. When the cue was incongruent, however, participants responded no slower than if there were no gaze cue at all. Therefore, while they use gaze to plan their motions, participants quickly recover from erroneous planning when the point of linguistic disambiguation is reached. This facilitation occurs even in the randomized section, when participants could not know ahead of time whether gaze would be congruent, incongruent, or absent. In short, current results suggest that there are scenarios in which adding eye gaze cues to a robot's behavior is a worthwhile investment: at best, it increases comprehension and efficiency, and at worst (when the gaze cue is in error), there is little damage to performance.

Other research has shown that incongruent gaze hinders performance in robot-instruction tasks (Staudte & Crocker, 2011). However, our study's task involves a lighter cognitive load, which may explain our divergent findings. In both studies, participants identify the referent of the robot's gaze and speech, but in our task, they simply select that referent, whereas in Staudte and Crocker's task, they compare features of that referent to features of other visible objects and then decide if a given statement is true or false. Thus, our experiment's task requires less cognitive processing, which may allow people to quickly overcome the incongruent gaze. This conjecture is supported by findings from a different study that used a similar task to ours (Huang & Mutlu, 2012). This study also found no difference between no gaze and incongruent gaze, while confirming the benefit of congruent gaze.

An alternate explanation is that the agent looks at the participant before speaking on incongruent trials, but not on no gaze trials, which may cue participants for the impending selection and negate any hindering effects of incongruent gaze. A revised no gaze condition in which the robot looks at the participant but not to a block would clarify this possibility.

Although mean response times did not significantly differ between robot and human agents, some differences did emerge between these conditions. Participants in the human agent group responded more quickly on average, although the difference was not significant (possibly because of the small group size). Perhaps relatedly, the error rate for participants in the human agent group (2.7%) was higher than the error rate for participants in the robot group (0.3%).

In response to the post-interaction survey question asking whether they noticed "anything unusual" during the experiment, five of the nine participants in the human agent group (56%) made reference to intentional misdirection by the actor, writing things like "She built up my trust and then betrayed me" and "She tried to trick me with her gaze." In comparison, only six of the 22 people in the robot group (27%) included such statements about intentional action from the robot. Even with identical behaviors, there was some difference in agency attributions between robots and humans.

However, human gaze is inherently less precise than robot gaze. Future experiments could record the human actor's face to verify that human gaze timings were comparable to robot timings. To generalize the results, future work should also test different collaborative scenarios to understand the conditions under which facilitation is possible without hindrance. Eye tracking would reveal at which point people decide to follow or ignore a robot's gaze. Future work should also randomize the assignment of conditions, rather than recruiting independent groups of participants, to rule out the possibility of group effects causing the observed variations.

MyKeepon has limited articulation and a simplified appearance. We chose this robot intentionally—MyKeepon's large eyes and gross body movements make its eye direction highly salient—but it is simpler and smaller than many other robots. Our results, therefore, may be most applicable to this type of robot. Future studies should investigate robots with articulated eyes as well as anthropomorphic robots to find whether physical appearance and eye motion affect gaze cues.

Acknowledgments

This work is supported by NSF grants 1139078 and 1117801. Thank you to Iulia Tamas and Terin Patel-Wilson for their assistance, and to reviewers for their helpful comments.

References

Andrist, S., Pejsa, T., Mutlu, B., & Gleicher, M. (2012). Designing Effective Gaze Mechanisms for Virtual Agents. In *ACM Annual Conference on Human Factors in Computing Systems (CHI 12).* Austin, Texas: ACM Press.

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, *3*, 41–52.

Boucher, J.-D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., et al. (2012). I Reach Faster When I See You Look: Gaze Effects in Human-Human and Human-Robot Face-to-Face Cooperation. *Frontiers in Neurorobotics*, *6*, 1–11.

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*(8), 1129–1138.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329.

Griffin, Z. M., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological Science*, *11*(4), 274–279.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596–615.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Huang, C.-M., & Mutlu, B. (2012). Robot Behavior Toolkit: Generating effective social behaviors for robots. In *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI '12).*

Huang, C.-M., & Thomaz, A. L. (2011). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2011).* Atlanta, GA USA.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100.

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon: A Playful Robot for Research, Therapy, and Entertainment. *International Journal of Social Robotics*, *1*, 3–18.

Michalowski, M., Machulis, K., & Gasson, M. (2013, July). *BeatBots MyKeepon GitHub Repository.* `https://github.com/beatbots/MyKeepon`.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In *6th IEEE-RAS International Conference on Humanoid Robots (Humanoids '06).*

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Footing in Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues. In *4th ACM/IEEE International Conference on Human Robot Interactions (HRI '09).* La Jolla, California: ACM.

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *4th ACM/IEEE International Conference on Human Robot Interactions (HRI '09).* La Jolla, California: ACM.

Pierno, A., Becchio, C., Wall, M., Smith, A., Turella, L., & Castiello, U. (2006). When gaze turns into grasp. *Journal of Cognitive Neuroscience*, *18*(12), 2130–2137.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*, 268–291.

Strabala, K., Lee, M. K., Dragan, A., Forlizzi, J., Srinivasa, S. S., Cakmak, M., et al. (2013). Toward Seamless Human-Robot Handovers. *Journal of Human-Robot Interaction*, *2*(1), 112–132.

Yu, C., Schermerhorn, P., & Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, *1*(2).