Do People Spontaneously Take a Robot's Visual Perspective?

Xuan Zhao
Brown University
190 Thayer Street
Providence, RI 02912, USA
1-(401)-863-2544
xuan_zhao@brown.edu

Corey Cusimano
Brown University
190 Thayer Street
Providence, RI 02912, USA
1-(401)-863-2544
corey_cusimano@brown.edu

Bertram F. Malle
Brown University
190 Thayer Street
Providence, RI 02912, USA
1-(401)-863-6820
bfmalle@brown.edu

ABSTRACT
This study takes a novel approach to the topic of perspective taking in HRI. In a human behavioral experiment, we examined whether and in what circumstances people spontaneously take a humanoid robot's visual perspective. We found that specific nonverbal behaviors displayed by a robot—namely, referential gaze and goal-directed reaching—led human viewers to take the robot's visual perspective, though marginally less frequently than when they encounter the same behaviors displayed by another human. This project identifies specific features of robot behavior that trigger spontaneous social-cognitive processes in human viewers and informs the design of interactive robots in the future.

Categories and Subject Descriptors
I.2.9 [**Artificial Intelligence**]: Robotics – operator interfaces. J.4 [**Computer Applications**]: Social And Behavioral Sciences – *Psychology*.

General Terms
Design, Experimentation, Human Factors.

Keywords
Human-robot interaction (HRI); communication; perspective taking; nonverbal behaviors; humanoid robot.

1. INTRODUCTION
Some of the challenges in human communication can be overcome by effective visual perspective taking (VPT)—by judging whether certain objects are visible to one's communication partner, and how they appear from the partner's specific viewpoint. A wealth of psychological research indicates that through VPT, partners identify shared knowledge, establish common ground, and resolve referential ambiguity [1]. What role might VPT play in human-robot interaction? Previous HRI research has focused primarily on enabling *robots* to take human partners' perspectives, exploring what architecture, strategies, and designs robots should acquire to handle possible perspective ambiguities [2]. Our current project, however, examines the human side: When encountering a humanoid robot, do people spontaneously attribute human-like properties, such as visual perspective, to robots? And if they do, in what circumstances are people more likely to take the robot's visual perspective? Finding answers to these questions will deepen our understanding of people's perception of robots as social agents; in addition, it will inform designers about the features and capacities robots should have to facilitate human-robot interaction.

Specifically, we examined the effect of an agent's nonverbal behaviors, such as referential gaze and goal-directed reaching, as triggers of spontaneous VPT in human viewers. Previous research indicated the important role of these behaviors in human-human interaction [3], and the same may be true in HRI, where nonverbal behaviors are of great relevance to designing social robots.

2. METHODS
2.1 Procedures
To capture people's spontaneous VPT, we created a single-trial task in which participants viewed a photograph depicting a human or robot sitting across a table. Placed on the table was a wooden digit "9", which was also interpretable as a "6" from across the table. While looking at this photo, participants answered the free-response question, "Which number is on the table?" A response of "6" indicated spontaneous perspective taking.

2.2 Design
We conducted a 2 × 4 between-subjects experiment. One manipulated factor was the type of agent participants encountered (human or humanoid robot). The humanoid robot was a 58-cm tall red-colored Nao (Aldebaran Robotics). Importantly, because Nao does not have white sclera (which plays an essential role in gaze following and joint attention in human interaction [4]), we edited its pupils in Adobe Photoshop CS6 to make its eyes indicative of gaze direction like human eyes do (see the left column in Figure 1 for an example of gaze).

The second manipulated factor was the behavior that the agent performed toward an ambiguous object. Participants saw the agent either 1) looking away from the object, thus being merely present in the scene (*presence*), 2) gazing at the object (*gaze*), or 3) reaching for while gazing at the object (*reaching*).

Besides three behavior conditions, we also created two control conditions. The novelty control measured people's spontaneous perspective taking when encountering a novel artifact "sitting" across a table (a colorful electric guitar). In the absolute baseline control, we determined the rate of spontaneous VPT with no agent or chair present in the scene.

HRI'15 Extended Abstracts, Mar 02-05 2015, Portland, OR, USA.
ACM 978-1-4503-3318-4/15/03.
http://dx.doi.org/10.1145/2701973.2702044

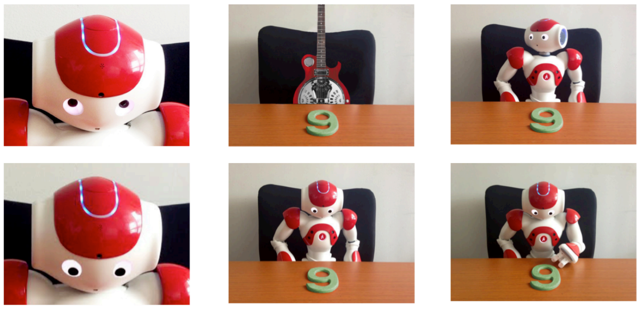

Figure 1. Nao, the humanoid robot in the current study. Left column: the original eyes (above) and photo-edited eyes (below). Middle and right columns: four photographs in humanoid robot condition: control, presence, gaze, and reaching (clockwise).

3. RESULTS

Of 365 participants (mean age = 33, 57% females) in the robot condition, 364 settled on one perspective and answered either “6” or “9”. Figure 2 depicts their VPT rates in each behavior condition (N = 86-92 per condition). A logit analysis with Helmert contrasts across four triggering conditions (three behavior conditions plus novelty condition) revealed that, compared with the VPT rate of the novelty condition (8.3%), the three behavior conditions induced significantly higher VPT rates, $z = 3.2$, $p < .01$. Among these behavior conditions, the average of gaze and reaching induced a significantly higher VPT rates than mere presence (16.3%), $z = 2.3$, $p < .05$; whereas rates in the gaze condition (28.6%) and the reaching condition (30.4%) were not significantly different from one another.

Of 235 participants (mean age = 30, 46% females) in the human condition (N = 57-63 per behavior condition), all answered either “6” or “9”. According to the logit analysis, compared with the absolute baseline condition (1.7%), the three behavior conditions again induced significantly higher VPT rates, $z = 3.4$, $p < .001$. In turn, compared with mere presence (12.7%), the average of gaze and reaching induced a significantly higher VPT rate, $z = 4.0$, $p < .001$, while the gaze condition (42.1%) and the reaching condition (45.6%) did not differ from one another.

An overall 2 (agent: robot, human) × 3 (nonverbal behavior: presence, gaze, reaching) logit analysis revealed no significant main effect of agent; however, the interaction between agent and the specific contrast of presence vs. gaze/reaching was marginally significant, $z = 1.7$, $p < .10$. That is, overall, robot and human agents triggered VPT in similar ways, but a human agent could trigger an even more powerful response in human viewers when showing gaze or reaching behavior, compared with being present.

4. DISCUSSION

This project is a first attempt to examine if people spontaneously take a robot’s visual perspective. We found that people indeed took a humanoid robot’s perspective, especially when it displayed nonverbal behaviors such as referential gaze and goal-directed reaching—which were the same triggering conditions that caused people to take another human’s point of view. In addition, people were more inclined to take another person’s visual perspective than a robot’s perspective when both agents exhibited the same nonverbal behaviors, but the difference was only marginal.

One limitation of our study is the use of static photographs. Although people are familiar with other humans’ nonverbal behaviors, many participants had not previously seen a robot perform these behaviors; they may therefore have had difficulty interpreting the robot’s behaviors as goal-directed. We are currently designing videos of Nao displaying the aforementioned behaviors in motion to investigate people’s perspective-taking tendency under these more dynamic conditions. Meanwhile, we are also investigating whether visual features of a robot (e.g. the presence of eyes and the overall appearance of the robot) have an impact on people’s perspective-taking tendency. To answer this question, we are currently replicating the same experiment design with Rethink Robotics’ Baxter, which looks much more masculine and whose screen can display eyes or not. Eyes are a critical (though perhaps not necessary) trigger of agency attributions, and it is an intriguing question whether cameras installed on a robot are enough to count as a *visual perspective*. We further plan to explore spontaneous perspective taking using more ecologically valid paradigms, where people collaborate with robots in joint tasks and verbally describe the objects to one another when taking different perspectives. Findings from such research will have implications on how humans view robots as social agents and how robots can be designed to achieve smoother human-robot interaction.

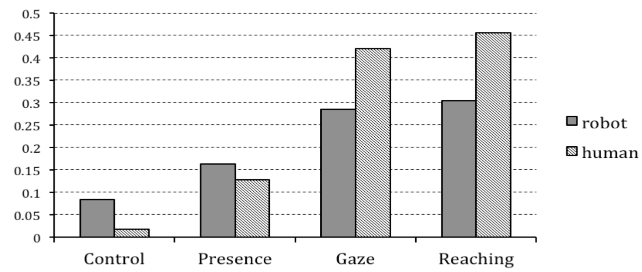

Figure 2. Spontaneous Visual Perspective Taking (VPT) rates for robot and human agents across triggering conditions

5. ACKNOWLEDGEMENT

This project was supported in part by a grant from the Office of Naval Research, No. N00014-14-l-0144. The opinions expressed here are our own and do not necessarily reflect the views of ONR. The authors are grateful for the support from Matthias Scheutz’s Human-Robot Interaction Laboratory at Tufts University, particularly Megan Strait.

6. REFERENCES

[1] Brennan, S. E., Galati, A., and Kuhlen, A. 2010. Two minds, one dialogue: Coordinating speaking and understanding. In *The Psychology of Learning and Motivation*, 53, 301–344. DOI= http://doi.acm.org/10.1016/S0079-7421(10)53008-1

[2] Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., and Schultz, A. C. 2005. Enabling effective human–robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 35, 4 (Jul. 2005), 460–470. DOI= http://doi.acm.org/10.1109/TSMCA.2005.850592

[3] Tversky, B., and Hard, B. M. 2009. Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110, 1 (Jan. 2009), 124–129. DOI= http://doi.acm.org/10.1016/j.cognition.2008.10.008

[4] Tomasello, M., Hare, B., Lehmann, H., and Call, J. 2007. Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *J. Hum. Evol*., 52, 3 (Mar. 2007), 314–20. DOI= http://doi.acm.org/10.1016/j.jhevol.2006.10.001