Neural Networks 23 (2010) 966–972

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2010 Special Issue

"Social" robots are psychological agents for infants: A test of gaze following

Andrew N. Meltzoff [a,*], Rechele Brooks [a], Aaron P. Shon [b], Rajesh P.N. Rao [b]

[a] Institute for Learning & Brain Sciences, University of Washington, Box 357920, Seattle, WA 98195, USA

[b] Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Keywords:
Social robots
Gaze following
Imitation
Social cognition
Bayesian learning

ABSTRACT

Gaze following is a key component of human social cognition. Gaze following directs attention to areas of high information value and accelerates social, causal, and cultural learning. An issue for both robotic and infant learning is whose gaze to follow. The hypothesis tested in this study is that infants use information derived from an entity's interactions with other agents as evidence about whether that entity is a perceiver. A robot was programmed so that it could engage in communicative, imitative exchanges with an adult experimenter. Infants who saw the robot act in this social-communicative fashion were more likely to follow its line of regard than those without such experience. Infants use prior experience with the robot's interactions as evidence that the robot is a psychological agent that can see. Infants want to look at what the robot is seeing, and thus shift their visual attention to the external target.

1. Introduction

Gaze following is a key component of human social cognition. Adults pay attention to the gaze of other people and are motivated to look at what others are seeing. *Homo sapiens* are more dependent upon observational learning than any other species, and one mechanism for acquiring information about people and things is monitoring others' looking behavior. Informational value is not homogenously distributed in the visual field. Where should one attend to maximize the likelihood of learning? Adults often visually inspect areas of space in which novel objects and events can be seen—informational hotspots. By following the gaze of others, one acquires input for social, causal, and cultural learning.

Gaze following contributes to the four key strands of social cognition. First, gaze following helps language learning; it assists in learning the names of things to follow the line of regard of the speaker who is labeling them. Second, gaze following helps learning about emotions; when an agent is surprised, fearful, or disgusted by a visual event, a learner can clarify what this emotion is about by following her line of regard. Third, gaze following contributes to imitative learning; goal-directed human acts are often visually guided, and a learner can hone in on goals and intentions by following gaze. Fourth, an individual's interests, desires, and stable preferences can be discerned by paying attention to that individual's looking patterns.

The role of gaze following in informal apprenticeship learning is amply shown in the foregoing examples, but gaze following is also instrumental in formal education and classroom settings. Didactic teaching makes frequent use of gaze to signal where learners should look in the classroom and to designate which student should respond. Typically developing adults and K-12 students effortlessly follow others' line of regard; however, children with autism spectrum disorders (ASD) have profound deficits in gaze following (Baron-Cohen, 1995; Mundy, Sullivan, & Mastergeorge, 2009; Toth, Munson, Meltzoff, & Dawson, 2006). This front-end deficit in children with ASD is detectable in infancy, persists in childhood, and cuts them off from many other forms of social learning in both informal and formal settings.

The realization that gaze following plays an important role in human social learning has sparked growing interest in endowing robots with the ability to follow human gaze. While some efforts have focused on imitating head movements (Demiris, Rougeaux, Hayes, Berthouze, & Kuniyoshi, 1997) and achieving gaze following for human–robot interaction through pre-programmed behaviors (Breazeal & Scassellati, 2000; Imai, Ono, & Ishiguro, 2001; Scassellati, 2002), others have explored methods for the development and emergence of gaze following (Breazeal, Buchsbaum, Gray, Gatenby & Blumberg, 2005; Carlson & Triesch, 2004; Fasel, Deák, Triesch, & Movellan, 2002; Nagai, Hosoda, Morita, & Asada, 2003). More recent efforts have proposed the use of probabilistic models for handling uncertainty and statistical learning to increase the accuracy of gaze following by adapting to a human's preferences during the course of human–robot interaction (Hoffman, Grimes, Shon, & Rao, 2006).

Critical issues for both robotic and child learning are: Who do you follow and under what conditions? The conditions supporting gaze following in children have been investigated by developmental psychologists, and five findings are particularly informative.

* Corresponding author. Tel.: +1 206 685 2045; fax: +1 206 221 6475.
E-mail address: Meltzoff@u.washington.edu (A.N. Meltzoff).

0893-6080/$ – see front matter © 2010 Published by Elsevier Ltd
doi:10.1016/j.neunet.2010.09.005

First, 12- to 18-month-old infants follow the gaze of a person who turns to look at lateral targets but refrain from doing so when that person's view is blocked by opaque screens, blindfolds, or eye closure (Brooks & Meltzoff, 2002; Butler, Caron, & Brooks, 2000; Dunphy-Lelii & Wellman, 2004; Moll & Tomasello, 2004). In other words, infants do not simply extend a line from the viewer's eyes and treat the gazer as "seeing through" such occluders. Second, gaze following is not an all-or-none skill, but depends on contextual factors such as viewing angles, distance and social context (Butterworth & Itakura, 2000; Butterworth & Jarrett, 1991; Collicott, Collins, & Moore, 2009; Deák, Flom, & Pick, 2000; Johnson, Ok, & Luo, 2007; Senju & Csibra, 2008). Third, gaze following changes with age (Brooks & Meltzoff, 2005; Carpenter, Nagell, & Tomasello, 1998; Moore & Povinelli, 2007; Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007). Fourth, gaze following is causally and developmentally linked to a network of other social-cognitive skills (Baldwin, 1993; Csibra & Gergely, 2009; Itakura et al., 2008; Johnson, 2003; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Repacholi, Meltzoff, & Olsen, 2008; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Fifth, infants' likelihood of gaze following is altered based on specific laboratory experience (Corkum & Moore, 1998; Meltzoff & Brooks, 2008).

There has been less work on the question of whom (or what) to follow. Although adults follow a person who re-orients his head towards an object in external space, they do not do so when a swivel chair spins or a teapot rotates about its vertical axis. What defines a perceiver worth following, and how does an infant recognize one?

These issues can be addressed by asking whether it is possible to design a robot that an infant will treat as a psychological agent that can see the external world. The hypothesis investigated in this paper is that infants use information derived from an entity's *interactions with other agents* as evidence about whether that entity is a perceiver. To test this in the present study, a commercially available robot was programmed so that it could engage in communicative, imitative games with an adult experimenter. After varying the infant's prior experiences with the robot in particular ways, infants were tested to see if they would follow the gaze of the robot to a distal target. Results showed that the nature of the prior experience mattered.

Importantly, in the current study the robot did not interact with the infant. The infant simply *observed* the adult–robot interactions as a bystander. The robot did not move contingently based on the infant's actions (as was done in two previous studies: Johnson, Slaughter, & Carey 1998; Movellan and Watson, 2002).

The study reported here is distinguished in three ways from previous studies using robots (and puppets) to investigate infant gaze following. First, in previous work the robot/puppet beeped, moved, or lit up in response to the infant's actions. The Movellan and Watson (2002) study was explicitly designed to ensure that infants learned that the robot was under their control. In the current work, infants simply witnessed how the robot interacted with a third party (see also, Johnson et al., 2008).

Second, in the Movellan and Watson (2002) study infants were trained by pairing acoustic and visual stimuli on the left with the robot's turning toward them. Thus there was explicit reinforcement connecting salient lateral events and the direction of the robot's turn, and this may have carried over to the test trials. In the present study, the first time the robot generated a lateral head turn was in the gaze-following test itself.

Third, in Johnson et al. (2008) infants watched an adult uncover a stuffed, oval-shaped object with no articulated parts or face. The object rotated laterally toward the adult, and the adult conversed with it while it beeped in response. Next the object rotated toward lateral targets. Infants looked in the direction that the object's front end turned; however, a lax measure of gaze following was used. There was no requirement that infants look at a distal target, only to look away from the agent and off midline—"either to the right or left of the midline anywhere in the vertical plane" (p. 28). The current study tested whether infants genuinely followed the entity's line of regard *to* the target object, not merely that the infant turned to the left or right of midline. Such referential specificity is needed to support socially mediated word, emotion, and causal learning.

The design of the current study entailed randomly assigning infants to one of four experimental groups. The *social-interaction* group saw adult–robot interaction. The robot engaged in communicative, imitative exchanges with the adult. For example, the adult raised her arm to a horizontal position and the robot imitated this act. The adult moved her arm to a vertical position, and the robot immediately did the same. The robot then took its turn in leading the imitative exchange. The robot twisted its arm, and the adult duplicated this act, and so on. To an adult observer this compelled a powerful "agentive illusion" that the robot could see what the adult was doing. The experimental question was whether this prior evidence changed the infant's subsequent tendency to follow the robot's gaze, to try to see what the robot was "seeing".

Three other groups were tested. In the *robot movement, passive adult* group, the robot performed the same actions in the same order as in experimental group but the adult did not respond. In the *robot–adult mismatch* group, the robot performed the same actions in the same order and the adult acted, but she did so in a way that mismatched the robot. In the *passive robot baseline* group, the experimenter performed the same actions as in the experimental group, but the robot remained passive. The adult movement was the same and the infant was familiarized with the robot's presence, but there was no evidence that the robot was a perceiver (and perhaps accumulating evidence that it was not).

The overarching question was: Does an infant's interpretation of the robot change as a function of observing the robot's behavior? The hypothesis tested was whether witnessing the robot's engagement in communication and imitation will bias infants to treat the robot as a psychological agent who can perceive, as measured by an increased likelihood of following the robot's "gaze".

2. Method

2.1. Participants

The participants were 64 18-month-old human infants ($M = 18.05$ months, ± 10 days of their 18-month-old birthday). Half of the participants were female. Infants were recruited by telephone from the University of Washington's infant participant pool, with the restriction that they be full-term, normal birth weight, and with no known developmental concerns, according to parental report. The racial/ethnic composition of the participants was 81.3% White, 4.7% Asian, 7.8% multiracial, with 6.2% of Hispanic ethnicity, according to parental report. The sample was primarily middle- to upper–middle class based on previous analyses of this university participant list. Additional infants were excluded from the study for the following reasons: robot mechanical problems ($n = 13$), extreme infant fussiness or lack of cooperation ($n = 14$), and experimenter error ($n = 3$).

2.2. Robot

The robot was a HOAP-2 humanoid robot manufactured by Fujitsu Laboratories, Japan. It was 50 cm tall and metallic in color. The robot had 25 degrees-of-freedom (DOF) including two hands that could be opened or closed, and a pan-tilt head with two black circles that were the rims of its miniature cameras (Fig. 1).

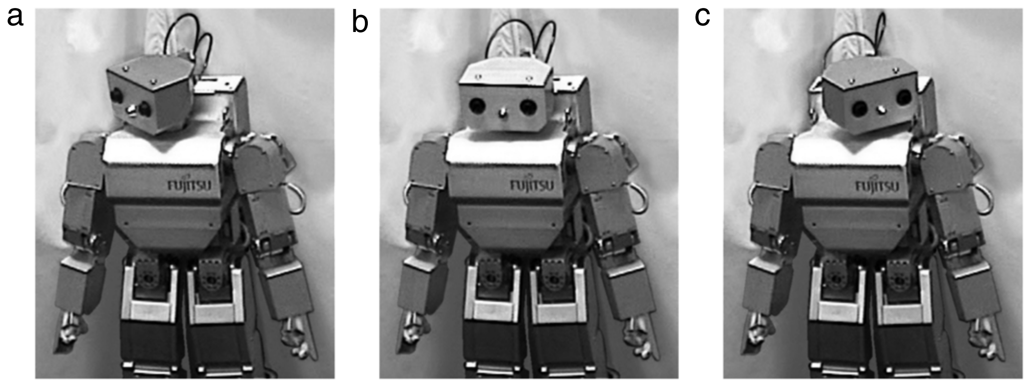

Fig. 1. Robot used to test gaze following in infants.

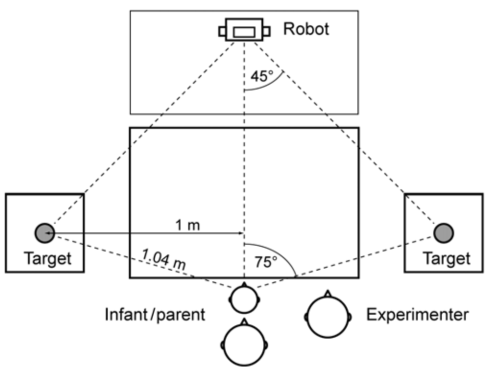

Fig. 2. Room diagram for robot–infant testing.

The robot stood on a platform (0.60 m high). There were two loudspeakers next to the robot's feet hidden from the infant. In an adjoining room, an assistant operated a computer program that controlled the robot's movements (hereafter, "operator").

2.3. Test environment and materials

Each infant was tested in a laboratory room (3.0 × 3.5 m) while seated on his or her parent's lap at a black table (1.14 × 0.76 m), directly facing the robot, which was 1.27 m away. The experimenter sat next to the infant, and used her arm closest to the infant to perform her gestures so the infant could easily monitor these movements (the side the experimenter sat on was counterbalanced within test groups and sex of participants). The room was lined with blue cloth, floor to ceiling, to provide a homogeneous visual background.

For the gaze-following test, two colorful, silent objects (9 cm diameter × 16 cm high) served as targets. These identical, plastic targets were placed on either side of the infant, and supported at eye level on top of pedestals. Each target was 75° off the infant's midline and 1.04 m away from the infant. The placement of the targets required the robot to make a 45° turn from midline to align with a target (Fig. 2).

Two digital cameras recorded the session, each on a separate DVD recorder. One focused on the infant's face and upper torso and the other recorded a frontal view of the robot's movements. A character generator added synchronized time codes (30 per s) onto both recordings, which were used for subsequent scoring from the digitized video record.

2.4. Design and procedure

Infants were randomly assigned to one of four groups: Group 1 (social interaction), Group 2 (robot movement, passive adult), Group 3 (robot–adult mismatch), and Group 4 (passive robot baseline). Sex of infant was counterbalanced within each group. To acclimate the infants, the experimenter played warm-up games for approximately 3 to 4 min. A white cardboard screen (0.51 m wide × 0.60 m tall) hid the robot during this time. During acclimation, the experimenter placed the target objects in the lateral positions (counterbalanced for first toy placed), talked to the infant, and played a game asking them about their nose and mouth.

After infants were acclimated, the experimental procedure involved two phases: exposure to a scripted robot experience and a subsequent gaze-following test. In Phase 1, infants watched as the experimenter and the robot engaged in scripted movements that differed according to the test group. In Phase 2, the robot turned to "look" at the lateral objects to assess infants' tendency to gaze follow.

2.4.1. Phase 1: scripted robot experience

Phase 1 began with the experimenter seated next to the infant. Two short beeps (0.3 s) were emitted from the loudspeakers. In response, the experimenter walked across the room, and removed the small, white screen to reveal the robot. As she returned to sit down next to the infant, another beep occurred. Following this, a different script was administered according to the infant's randomly assigned group. Each script consisted of 10 bouts of activity involving the robot and/or experimenter. The mean duration of the infants' exposure to the robot in Phase 1, from screen removal (revealing the robot) to the onset of attention-centering stimuli before the gaze-following test, was 1.67 min, with no significant difference among the groups, $F(3, 60) = 0.34$, $p = 0.80$.

The robot's movements were controlled by a graphical user interface (GUI, Fig. 3) implemented within the X Windows system in Linux. The GUI (a custom C++ program) allowed the operator to control the robot using either clickable buttons to initiate individual actions, or to load pre-recorded files containing a series of recorded movements. Each file consisted of a series of operations played back to the infant (e.g., robot executes "raise left arm 90°"). Each operation within the file was annotated with a time to initiate that operation (relative to the start of the series of operations). Script timing granularity was at the level of milliseconds.

2.4.1.1. Group 1 (social interaction). For the social-interaction group, the actions were arranged to give the impression of a give-and-take communicative interaction between the adult and the robot. The infant simply witnessed the human–robot social interaction as a bystander. The human operator in the adjacent room viewed the test session and triggered the robot's movements via the GUI according to the schedule specified in Table 1. Each of the robot's movements in the 10 pre-determined bouts of activity

Table 1
The script for the social-interaction group using A, B, C to indicate the order of the robot's and experimenter's behavior within each bout.

Bout	Robot	Experimenter
1	A. Wave arm from elbow with hand oriented vertically	B. Wave hand from elbow and say, "Oh. Hi! That's our robot"
2	B. Nod head (as if nodding yes)	A. "Do you want to play?"
3	B. Touch torso with one hand	A. "Where is your tummy?" C. "Here is my tummy" as touched torso
4	B. Touch head with one hand	A. "Where is your head?" C. "Here is my head" as touched head
5	B. Nod head (as if nodding yes)	A. "Do you want to play another game?"
6	B. Raise mirroring arm to horizontal position	A. "Can you do this?" as raising arm to horizontal position
7	B. Raise mirroring arm to vertical position	A. "Can you do this?" as raising arm to vertical position
8	B. Nod head (as if nodding yes)	A. "Do you want a turn?"
9	A. Wave arm from elbow with hand oriented down ("inverted wave")	B. Imitate the inverted wave with mirroring arm
10	A. Raise arm to horizontal pointing to infant, twist it back and forth from shoulder ("arm twist")	B. Imitate arm twist with mirroring arm

Fig. 3. Graphical User Interface (GUI) used to control the robot.

was triggered by the operator, resulting in a well-timed interaction between the adult and the robot. A key feature of the experiment is that all of the commands to the robot were stored in a computer file, so that the pace and order of the acts could be used for the other experimental groups following a yoked-control design.

Table 1 provides a verbal description of the social-interaction script; Fig. 4 shows representative snapshot frames of the robot executing different individual operations. In Bout 1 the robot waved, and the experimenter responded. In Bouts 2–5 the experimenter asked the robot questions, and it respond appropriately (e.g., "Where is your tummy?" was followed by the robot touching its torso). In Bouts 6 and 7 the experimenter showed the robot actions with her arm and the robot imitated. In Bout 8 the experimenter asked the robot if it wanted a turn demonstrating an action, and it nodded affirmatively. In Bouts 9 and 10 the robot demonstrated novel arm movements to the experimenter, and she imitated.

Throughout the script, the experimenter acted as if she was engaged in a normal social interaction with the robot and directed her speech and actions towards it. The illusion of natural human–robot communication was achieved: In pilot work, naïve adults (undergraduates and lab personnel who were naïve to the operator in the adjacent room) reported that they thought the robot "had acoustic and visual sensors that allowed it to be responsive to the adult's movements".

2.4.1.2. Group 2 (robot movement, passive adult). This group equates the infant's exposure to the robotic movements using a yoked-control design. For each infant in Group 1 (social-interaction group), a data file preserved the robot's operations, which were played back to individual infants in Group 2. Thus, the robot's movements were identical in Groups 1 and 2, the difference between groups being that in Group 2 the experimenter remained stationary. The robot appeared to be self-actuated but not engaged in communication.

2.4.1.3. Group 3 (robot–adult mismatch). This group uses the same robot and experimenter movements as in the social-interaction group, but disrupts the contingencies and social-communication in a different way from Group 2. Using the yoked-control method, the robot performed the same movements in the same order (1, 2, 3, . . . , 10, see Table 1) as in the social-interaction group, but the experimenter performed her actions in the reverse order (10, 9, 8, . . . 1, see Table 1). The impression to an adult observer was that the experimenter and the robot were not communicating, because the timing and form of the robot's actions and the experimenter's actions were not in synchrony. Importantly, the robot movements for Group 3 were identical to those in Groups 1 and 2.

2.4.1.4. Group 4 (passive robot baseline). This group provides a baseline measure of how infants respond in the subsequent gaze-following test without prior evidence about the robot's capacity for social interactions. In this group, the robot remained stationary. The experimenter executed her actions from memory following the order and timing of the social-interaction group as closely as humanly possible, but without any response from the robot.

For all groups, at the end of the script, the experimenter left the room, saying good-bye to the infant. The robot remained facing the infant.

2.4.2. Phase 2: gaze-following test

After the experimenter had left the room, the robot turned its head 45° toward one of the two lateral targets, which constituted a trial. The robot's head turn took 0.5 s and the robot faced the object for 6 s. This speed of head turn and 6.5 s trial format was modeled after studies of human gaze following (e.g., Brooks & Meltzoff, 2002; Meltzoff & Brooks, 2008). The direction of the head turns was either L, R, R, L or R, L, L, R (counterbalanced within groups and sex of participant, where L indicates left).

In order to attract the infant's attention to the robot before each test trial, hidden speakers emitted a brief tone (0.3 s) and the robot moved its head downward (15°) and back to a position that resembled looking at the infant. The beep and/or head movement were controlled by the operator via the GUI and were repeated if necessary to attract the infants' attention (as is typical in tests involving humans who speak and perform attention-getting acts before testing gaze-following). A test trial was launched when the infant was fixated on the robot's face. In rare instances, the infant looked down or away just as the trial started; those trials

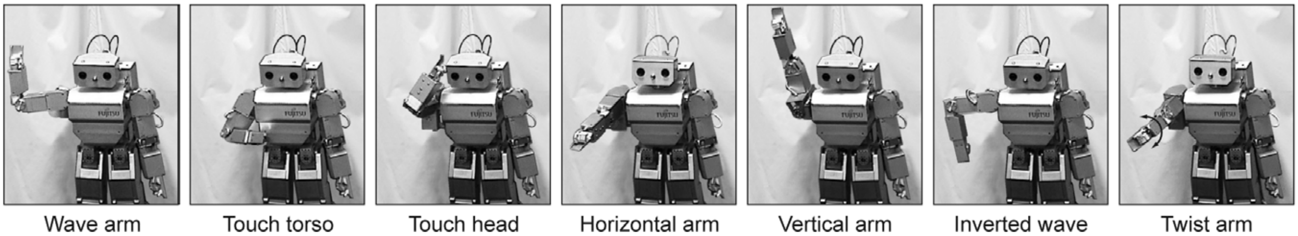

Fig. 4. Snapshots of robot movements.

were excluded as mistrials (3 mistrials in 256 trials, based on 64 participants × 4 trials each). As needed across the four test trials, the experimenter occasionally returned to the test room to help return the infant to the center of the table; in these cases, the experimenter left before test trials resumed. In Groups 1–3, the robot performed two arm movements before the first and third trial to attract attention to itself (arms bilaterally raised and lowered); this was not done in Group 4 because the robot, per experimental design, did not display any limb movements in this group.

2.5. Scoring of behavior

The video records of the infant's face were scored in a random order by an independent coder. The coder was naive to the infant's experimental group, and the direction of the robot's head turns. There was no record of the robot's head turns on the infant's video. The coder scored infants' looks at the targets. The operational definition of "look at target" was that the infant turned his head and aligned his eyes with the target location for at least 0.33 s (10 video frames). This is the same criterion commonly used in work on infant gaze following of human agents (e.g., Brooks & Meltzoff, 2002).

For each trial, an infant's first target look was designated as either a correct look if it matched the target the robot looked at (+1), an incorrect look if it was at the opposite target (−1), or a nonlook if the infant did not look at either target (0). Across the four trials, a total looking score measure was tallied, which ranged from −4 to +4. In addition, a second dependent measure was used, which assigned a dichotomous pass/fail score. This measure classified infants into those who followed the robot's gaze to the target on any trial ("pass") or those who did not ("fail").

Scoring agreement was assessed by rescoring a randomly selected 25% of the infants. The second coder was also kept blind to each infant's experimental group and direction of the robot's head turn. There were no intra- or interscorer disagreements, resulting in Cohen's kappas of 1.00.

3. Results

Infants followed the gaze of the robot as illustrated in Fig. 5, and as expected, looking scores varied as a function of experimental group. The looking scores were at their maximum for Group 1 (social interaction) ($M = 0.56$, $SD = 0.81$), and showed a linear decline for each of the other groups: Group 2 (robot movement, passive adult) ($M = 0.38$, $SD = 1.09$), Group 3 (robot–adult mismatch) ($M = 0.06$, $SD = 1.34$), and Group 4 (passive robot baseline) ($M = -0.31$, $SD = 0.79$). A one-way analysis of variance (ANOVA) with a test for a linear trend was statistically significant, $F(3, 60) = 6.47$, $p = .014$, $\eta_p^2 = 0.11$. Group 1 (social interaction) was the only group for which the mean looking scores were significantly greater than 0, $t(15) = 2.76$, $p = 0.015$, $d = 0.69$. The t-values for the Groups 2–4 were respectively 1.38, 0.19, and −1.58, $p > 0.10$ in all cases.

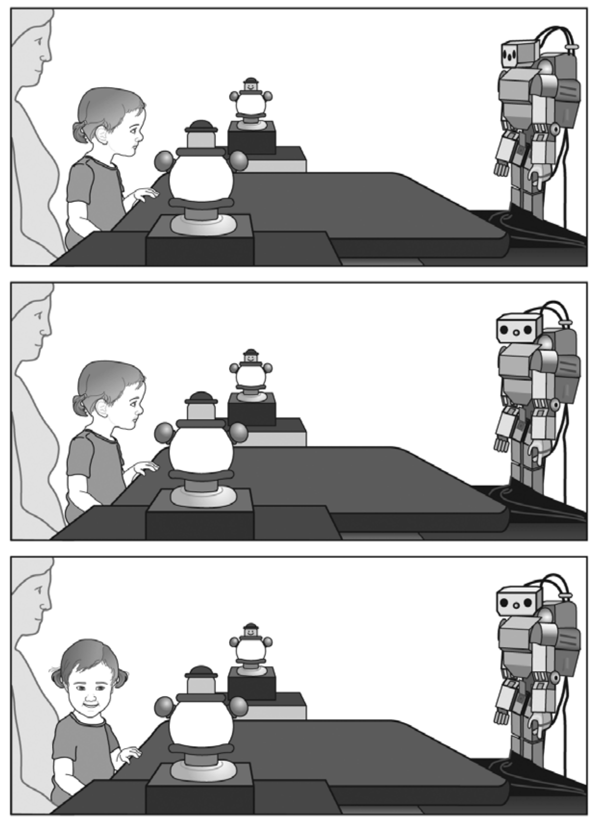

Fig. 5. Illustration of an 18-month-old infant following the gaze of a robot.

Table 2
Infant gaze following of robot as a function of prior experience with the robot.

Experimental group	Gaze following	
	Pass	Fail
Social interaction	13	3
Robot movement, passive adult	7	9
Robot–adult mismatch	9	7
Passive robot baseline	3	13

Table 2 displays the results for the dichotomous dependent measure that categorized infants according to whether or not they looked where the robot looked in any trial. The results show that more infants in Group 1 (social interaction) followed the robot's gaze (81.25%) than did so in Group 2 (43.75%), Group 3 (56.25%), or Group 4 (18.75%). A 4 (group) × 2 (pass/fail) chi-square test yielded a highly significant effect, $\chi^2(3, N = 64) = 13.00$, $p = 0.005$, Cramer's $\phi = 0.45$.

4. Discussion

The results show that infants' likelihood of following the "gaze" of a robot is influenced by their prior experience. Infants who see the robot act in a social-communicative fashion are more likely to follow its line of regard to an external target. Infants acquired information about the robot simply by observing its movements and interactions with the adult. The robot did not respond contingently to the infant's behavior. The infant was a bystander, eavesdropping on the social exchange between experimenter and robot. Moreover, the first time the robot rotated its head to a lateral target was during the gaze-following test. Robotic lateral head turns were excluded from prior experience. We interpret these findings as indicating that infants use their prior experience with the robot's actions and interactions as evidence that the robot is a psychological agent that can see specific targets in the external world.

This interpretation is compatible with Johnson et al. (1998, 2008) and Movellan and Watson (2002). We are currently investigating whether such capacities may provide a stepping-stone toward greater social ability, such as identifying a useful teacher from among several external agents in the learner's environment (Kaipa, Bongard, & Meltzoff, 2010).

Several alternative accounts of these findings can be rejected, given the experimental design and procedures. First, the dependent measure was whether infants shifted their visual gaze outward in space to the target specified by the robot's line of regard. It is not just that infants treated the robot's movement as a visually salient event. Had they done so, they may have watched the rotating cuboid atop the metallic body (the robot's "head") but not have extended their look to the distal target. The infants did not just track a corner of the moving cuboid, but directed their gaze to a distal target more than a meter away.

Second, the robot had humanoid physical features, but this alone did not drive the results. It is not sufficient that infants treat the robot's black disks as "eyes" and then generalize from human experience based on physical features alone. The robot was morphologically identical in all groups, yet infants' responses varied.

Third, the robot appeared to be a self-actuated agent that moved by itself, but this cannot account for the results across the groups. The robot in Groups 1–3 performed identical movements (via computer playback). This suggests that it is not solely the robots actions, but the nature of the *interaction* that carries weight.

What key aspects of the interaction made a difference? Who (or what) is worthy of driving your own visual system so that it aligns with this other entity? A novel proposition suggested by our results is that the robot's capacity for generative imitation is a powerful cue to psychological agency and communication for infants. The current experiment did not independently manipulate all possible factors, but we speculate that imitation acts as an especially salient cue to psychological agency in human infants. An entity that imitates is one whose "gaze" should be followed. An entity that imitates, takes turns, and is responsive to language provides even stronger evidence of being a psychological agent; future studies will need to determine the weights given to each.

In the social-interaction group, the robot imitated a variety of adult gestures. The adult performed acts on a random schedule and the robot copied them without being touched. How can such imitation at a distance be explained except through the presence of a perceptual apparatus that senses the visual input? The structural congruence manifest in imitation goes beyond simple temporal contingency detection alone. One could respond in a temporally contingent fashion, but with a behavioral mismatch which is not imitation. In a study designed to tease apart imitation from temporal contingency detection, Meltzoff (2007) showed that infants are sensitive to imitation over and above a temporal contingency.

It is noteworthy that gaze following was not an all-or-none deterministic process. Individual infants did not always look where the robot "looked". Furthermore, the probability of looking depended on the type of evidence accumulated during a period of prior observation of the robot's actions. This suggests that looking or not looking is governed by a stochastic decision-making process that could be modeled using a Bayesian approach. We hypothesize that infants use a model-based, probabilistic strategy for determining whether to follow an external actor's line of regard. We propose that the model used by the infant is derived from an internally estimated belief in the "psychological agency" of the entity. Indeed, infants may be accumulating evidence not only that the entity can see, but also that it can*not* see. The looking scores in the passive robot baseline group were the lowest of all groups. Infants may have taken data that the adult was speaking to the robot and showing it gestures – and the robot did nothing to respond – as evidence that the robot was not a psychological agent with sensors.

Bayesian models have been useful in illuminating causal, imitative, and gaze-following learning in children and robots (Gopnik et al., 2004; Gopnik & Schulz, 2007; Gopnik & Tenenbaum, 2007; Movellan & Watson, 2002; Rao, Shon, & Meltzoff, 2007; Shon, 2007; Shon, Storz, Meltzoff, & Rao, 2007; Tanaka, Cicourel, & Movellan, 2007; Verma & Rao, 2006). The results presented here suggest that infants are using sequential Bayesian inference to estimate the psychological agency of the entities they encounter based on a history of observed behaviors by those entities. In the sequential Bayesian inference model, the infant starts with a prior belief about the robot's agency based on past observations and interactions with people and robot-like toys. The infant then updates, in a Bayesian manner, the belief in the robot's agency on the basis of new evidence obtained from their own and others' continued interactions with the robot. One method to test this in future work would be to alter the robot's behaviors over time in a short-term longitudinal study to carefully track the extent to which a sequential evidence-accumulation process may or may not be occurring.

The present study demonstrates that social-communicative interaction plays a key role in mediating infants' gaze following of a robot. The robot's humanoid appearance alone was not sufficient to cause gaze following. This finding has implications for the future design of humanoid robots and for the field of social robotics in general (see Kuhl, 2010 on language learning). Seamless human–robot interaction requires not only that the robot be able to follow a human partner's gaze but also that the human is motivated to follow the robot's. Our results suggest that it may be helpful for the robot first to engage in some form of socially intelligent behavior – either with the human partner or with other persons – to increase the human observer's belief in its agency and thereby enhance gaze following.

In the present study, infants' observation that the robot socially interacted with an adult and imitated at a distance changed infants' interpretations of the heap of metal on the other side of the table. The swivel of the cuboid on top of the metallic entity was no longer interpreted as a random physical movement, but as a meaningful action—a perceptual act directed to the external target. Infants treated the robot as a psychological agent that could see.

Acknowledgements

This research was supported by grants from NSF (SBE-0354453 and grant #0622252) and ONR (N000140910097). The views expressed are the authors' and do not necessarily represent the views of the agencies. We thank Susan Johnson, Javier Movellan, Cynthia Breazeal, Rawichote Chalodhorn, David Grimes, Deepak Verma, and the LIFE Science of Learning Center group for useful discussions. Josh Bongard and Patricia Kuhl provided helpful suggestions on an earlier draft of this manuscript. Craig Harris, Calle Fisher, Tiffany Bybee, and Matt Hoffman assisted with the empirical work.

References

Baldwin, D. A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*, 832–843.

Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning from and about others: towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, *11*, 31–62.

Breazeal, C., & Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, *8*, 49–74.

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology*, *38*, 958–966.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535–543.

Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, *1*, 359–377.

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, *18*, 25–50.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55–72.

Carlson, E., & Triesch, J. (2004). A computational model of the emergence of gaze following. In H. Bowman, & C. Labiouse (Eds.), *Proceedings of the 8th neural computation and psychology workshop* (pp. 105–114). Singapore: World Scientific.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4, Serial No. 255).

Collicott, C., Collins, S., & Moore, C. (2009). The development of gaze following in a third-party context. *Infancy*, *14*, 363–376.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*, 28–38.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153.

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, *36*, 511–523.

Demiris, J., Rougeaux, S., Hayes, G. M., Berthouze, L., & Kuniyoshi, Y. (1997). Deferred imitation of human head movements by an active stereo vision head. In *Proceedings of the sixth IEEE international workshop on robot and human communication* (pp. 88–93). Los Alamitos, CA: IEEE Computer Society.

Dunphy-Lelii, S., & Wellman, H. M. (2004). Infants' understanding of occlusion of others' line-of-sight: implications for an emerging theory of mind. *European Journal of Developmental Psychology*, *1*, 49–66.

Fasel, I., Deák, G. O., Triesch, J., & Movellan, J. R. (2002). Combining embodied models and empirical research for understanding the development of shared attention. In *Proceedings of the 2nd international conference on development and learning* (pp. 21–27). Los Alamitos, CA: IEEE Computer Society.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological Review*, *111*, 3–32.

Gopnik, A., & Schulz, L. (2007). *Causal learning: psychology, philosophy, and computation*. New York: Oxford University Press.

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, *10*, 281–287.

Hoffman, M. W., Grimes, D. B., Shon, A. P., & Rao, R. P. N. (2006). A probabilistic model of gaze imitation and shared attention. *Neural Networks*, *19*, 299–310.

Imai, M., Ono, T., & Ishiguro, H. (2001). Physical relation and expression: joint attention for human–robot interaction. In *Proceedings of the 10th IEEE international workshop on robot and human interactive communication* (pp. 512–517). Los Alamitos, CA: IEEE Computer Society.

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: infants' imitation of a robot's goal-directed actions. *Infancy*, *13*, 519–532.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*, 549–559.

Johnson, S. C., Bolz, M., Carter, E., Mandsanger, J., Teichner, A., & Zettler, P. (2008). Calculating the attentional orientation of an unfamiliar agent in infancy. *Cognitive Development*, *23*, 24–37.

Johnson, S. C., Ok, S.-J., & Luo, Y. (2007). The attribution of attention: 9-month-olds' interpretation of gaze as goal-directed action. *Developmental Science*, *10*, 530–537.

Johnson, S. C., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science*, *1*, 233–238.

Kaipa, K. N., Bongard, J. C., & Meltzoff, A. N. (2010). Self discovery enables robot social cognition: are you my teacher? *Neural Networks*, *23*, 1113–1124.

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*, 713–727.

Meltzoff, A. N. (2007). 'Like me': a foundation for social cognition. *Developmental Science*, *10*, 126–134.

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: a training study in social cognition. *Developmental Psychology*, *44*, 1257–1265.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*, 284–288.

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*, F1–F9.

Moore, C., & Povinelli, D. J. (2007). Differences in how 12- and 24-month-olds interpret the gaze of adults. *Infancy*, *11*, 215–231.

Movellan, J. R., & Watson, J. S. (2002). The development of gaze following as a Bayesian systems identification problem. In *Proceedings of the 2nd international conference on development and learning* (pp. 34–40). Los Alamitos, CA: IEEE Computer Society.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*, 938–954.

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, *2*, 2–21.

Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*, *15*, 211–229.

Rao, R. P. N., Shon, A. P., & Meltzoff, A. N. (2007). A Bayesian model of imitation in infants and robots. In C. L. Nehaniv, & K. Dautenhahn (Eds.), *Imitation and social learning in robots, humans, and animals: behavioural, social and communicative dimensions* (pp. 217–247). Cambridge: Cambridge University Press.

Repacholi, B. M., Meltzoff, A. N., & Olsen, B. (2008). Infants' understanding of the link between visual perception and emotion: 'if she can't see me doing it, she won't get angry'. *Developmental Psychology*, *44*, 561–574.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*, 13–24.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*, 668–671.

Shon, A.P. (2007). *Bayesian cognitive models for imitation*. Unpublished doctoral dissertation. Seattle, WA: Department of Computer Science and Engineering, University of Washington.

Shon, A. P., Storz, J. J., Meltzoff, A. N., & Rao, R. P. N. (2007). A cognitive model of imitative development in humans and machines. *International Journal of Humanoid Robotics*, *4*, 387–406.

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, *104*, 17954–17958.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Toth, K., Munson, J., Meltzoff, A. N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders*, *36*, 993–1005.

Verma, D., & Rao, R. P. N. (2006). Goal-based imitation as probabilistic inference over graphical models. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems, 18* (pp. 1393–1400). Cambridge, MA: MIT Press.