Mentalizing to Non-human Agents by Children

Shoji Itakura (sitakura@bun.kyoto-u.ac.jp)
Kyoto University & ATR Intelligent Robotics and Communication Lab.

Takayuki Kanda (kanda@atr.jp)
ATR Intelligent Robotics and Communication Lab.

Hiroshi Ishiguro (ishiguro@ams.eng.osaka-u.ac.jp)
Osaka University & ATR Intelligent Robotics and Communication Lab.

Abstract

In Experiment 1, fifty two-year-old children were tested to examine whether they could reproduce the target outcome of a robot in a goal re-enactment paradigm developed by Meltzoff (1995). The results show that the children were not only able to reproduce the target action produced by the robot, but were also able to complete the same task which the robot attempted, but failed to perform. However, it was essential that the robot mimicked human behavior suggesting intention, such as gazing at a partner and at the object being manipulated, in order to induce children to produce the target outcome in the failed attempt condition. In Experiment 2, a standard False Belief Task was conducted with a robot to investigate whether preschoolers attribute false belief to a robot or not. Results suggested that the children attribute false belief to a robot but don't attribute a mental verb to it.

Introduction

When do children first attribute mental states to others, and when they do, to whom do they attribute the mental state? Several studies have suggested that children comprehend goal-directed behavior from an early age (Carpenter et al., 1998; Csibra, 2003; Frith & Frith, 2003; Gergely et al., 1995). Woodward (1998) developed a new paradigm for understanding goal-directedness using visual habituation. She tested whether infants encode human action in terms of an actor's goals or in terms of spatiotemporal movement. In her experiments, infants viewed a hand reaching towards one of two objects. Upon habituation, the location of these objects was switched and the experimenter reached either towards the other object in the same location or the original object in its new location. It was observed that both five- and nine-month-old infants looked for longer when the experimenter reached towards the new object than when she grasped the old object in its new position. Woodward therefore concluded that young children tend to encode the actions of other people as goal-directed. These results, and those of subsequent studies, suggest that infants attribute an intentional relationship between objects and the world (Johnson, 2000). Meltzoff (1995) produced further evidence of goal comprehension in infants using the re-enactment of goals paradigm. In his study, 18-month-old infants reproduced the aims of the object-directed actions of adults, even in cases when the goals set within the model were never actually attained, but had to be inferred. However, under conditions in which the human agent was replaced by mechanical pincers performing the same actions, infants did not achieve the unattained goals. Meltzoff concluded that whereas 18-month-old infants were able to gauge the intentions of a human and complete the failed action, this was not the case with a mechanical pincer, to which they did not attribute goals. Johnson, Booth, and O'Hearn (2001) studied infant imitation and the production of communicative gestures, starting from the hypothesis that the recognition of mentalistic agents is not isomorphic with person recognition. Rather, it is based on non-arbitrary object perception, including the presence of a face and the ability to interact contingently with other agents. These authors replicated Meltzoff's study but modified it by using a stuffed orangutan as the non-human agent. They found that 15-month-old infants re-enacted the goals of an inanimate object that had a face and interacted contingently with the infants and the experimenter. In Experiment 1, adopting the same perspective as that of Johnson, Booth, and O'Hearn (2001), and using Meltzoff's (1995) re-enactment of goals paradigm, we investigated firstly whether young children imitate the actions of an autonomous humanoid robot, and secondly, whether they attain the goals indicated by its incomplete action.

One of the most important milestones in social cognitive development is to understand another's false belief. The False Belief Task developed by Wimmer and Perner (1983), also known as the "Maxi Task," measured this ability. It goes as follows: Maxi has some chocolate and puts it into a blue cupboard. Then he goes out. Now his mother comes in and moves the chocolate to a green cupboard. Maxi comes back to get his chocolate. Where will Maxi look for the chocolate? The answer is Maxi will look in the blue cupboard, because this is where he erroneously believes the chocolate to be. A series of studies established that children of around four years old begin to pass this task and can verbally explain it when asked.

In Experiment 2, we conducted the False Belief Task with a robot to establish whether preschoolers attribute false belief to a robot or not.

Experiment 1: Inference of a robot's goal by young children

Method

Participants: The sample consisted of 50 children (24 boys and 26 girls). Seven additional infants were excluded because they failed to complete all the test trials. Participants were all aged between 24 and 35 months (M=30.6, SD=3.2).

Stimuli: The experiment took place in an infant laboratory at Kyoto University. We employed an autonomous robot named Robovie, developed at the ATR Intelligence Robotics Laboratory in Kyoto, Japan. Robovie is an autonomous humanoid robot (1.2 m tall, with a 50 cm radius, and weighing 40 kg) that can move independently, and has human-like eyes and hands. It is equipped with visual, auditory, and tactile sensors, designed to enable it to imitate human behavior. Robovie can engage in communicative behavior with humans and mimics human behavior such as shaking hands, joint visual attention, and pointing.

In the present study, unlike the experiments of Meltzoff (1995) and Johnson, Booth, and O'Hearn (2001), the agent's action was presented to the children on a video monitor (38 x 64 cm). We considered it reasonable to use a video monitor to present these stimuli, as Barr and Hayne (2000) previously reported that 18-month-old infants imitated target actions in a video monitor condition. Phillipes and Wellman (in press) have also demonstrated the validity of using videotaped actions in research with infants. There were two action trials, a full-demonstration (complete) action, and a failed-attempt (incomplete) action.

Each action (complete, incomplete) trial consisted of two gaze conditions according to the robot's gazing behavior in relation to a human social partner (Figure 1). Thus there were four types of video stimuli and a baseline condition. Each video stimulus lasted 20 seconds, and included the successive manipulation of three different sets of objects.

1) Full Demonstration + Gaze: The infant watched the robot act on each set of objects three times successfully. The robot gazed at its partner's face before beginning each task, then looked at the object while manipulating it, and finally gazed at its partner's face again after completing each set of actions.

2) Full Demonstration + No Gaze: The subject watched the robot act on each set of objects three times successfully, but unlike in the gaze condition, the robot kept looking forward during the task.

3) Failed Attempt + Gaze: The subject watched the robot act unsuccessfully on each set of objects three times. The robot gazed at its partner's face before beginning each task, then looked at the object while manipulating it, and finally gazed at the partner's face again after failing to complete each task.

4) Failed Attempt + No Gaze: The subject watched the robot act on each set of objects three times unsuccessfully, but unlike in the gaze condition, the robot kept looking forward during the task.

5) Baseline: In the baseline condition, each trial consisted of the child manipulating the object for 20 seconds without visual stimuli.

There were three sets of objects: a dumbbell, a cup and beads, and a peg with an elastic band.

The dumbbell. In the complete condition, the experimenter handed the object to the robot, which grasped one end of the dumbbell in each hand, pulling the two ends apart. For the incomplete condition, the robot grasped the dumbbell in the same manner, but one hand slipped off the end of the dumbbell before it came apart.

The cup and beads. In the complete condition, the experimenter handed the beads to the robot with the string above the edge of the cup and the robot subsequently dropped the beads inside the cup. For the incomplete condition, the robot grasped the beads, lifted the string above the edge of the cup, wavered slightly over it, and then dropped the beads outside the cup.

The peg and elastic band. In the complete condition, the experimenter handed the robot an elastic band, which it grasped and hung on the peg. For the incomplete condition, the robot grasped the elastic band, raised it up towards the peg, but released it just before it circled the peg, thus dropping it onto the table.

Procedure: During each session, the child was seated in front of a small table facing a video monitor, with his/her parent or caregiver seated behind or next to them. After a five-minute habituation period, the experimenter began operating the video monitor for the presentation of the stimuli. The upper half of the infant's body was monitored by a video camera placed under the video monitor.

After they had viewed the video stimuli, the object they had just seen the robot manipulate was placed in front of the child by the experimenter. The sequence of the three objects was fixed, as the order of presentation was not found to have a significant effect in previous studies (Meltzoff, 1995; Johnson et al., 2001). If the child did not touch the objects, the experimenter would call its name or say "Look!" to engage his/her attention, but did not give any direct instructions. The experimenter gave neither affective nor linguistic cues during the viewing of the video stimulus and the response period.

Results and discussion

Since there were three target actions, the score achieved by each infant ranged from 0–3. A child obtained 3 points if he/she completed the target action with all three object sets, and if he/she failed to complete the goal using any of the sets his/her score was 0. The mean score for the *Full Demonstration + Gaze* condition was 2.2; 2.1 for the *Full Demonstration + No Gaze* condition; 1.6 for the *Failed Attempt + Gaze* condition; 0.7 for the *Failed Attempt + No Gaze* condition, and 0.4 for the *Baseline Condition*. The resulting overall mean value for each condition is shown in Figure 2. An analysis of variance (ANOVA) was performed on the effect of the *Gaze* condition on the *Full Demonstration*/*Failed Attempt* condition. A significant effect of the *Gaze* condition [$F(1, 36) = 4.29$, $P < 0.05$] on the *Full Demonstration*/*Failed Attempt* condition [$F(1, 36) = 17.14$, $P < 0.001$] was found. The interaction between the *Gaze* condition and the *Full Demonstration/Failed Attempt* condition was not significant [$F(1,36)=2.74$, $p < 0.106$]. Only the *Failed Attempt* + *No Gaze* condition was not significantly different from the baseline condition [$t(18)=0.878$, n.s.].

No difference was observed in the children's performance, irrespective of whether or not the robot gazed at its partner's face during the full demonstration; the children imitated the robot's actions. The failed attempt, in which the children observed the robot's attempt and failure to produce the target outcome, was the most interesting. In this condition, the children produced the target outcomes when the robot looked at the partner and the object; however, they failed to produce the intended action when it did not exhibit such intention-implying behavior. In the baseline condition, the children rarely produced the target outcomes; this result is consistent with those obtained by Meltzoff (1995) and Johnson, Booth, and O'Hearn (2001).

Infants are known to distinguish between humans and inanimate objects. By two months of age, children treat people as social entities, smiling, vocalizing, and imitating their actions, but objects are treated as toys to be looked at and to be manipulated (Legerstee, 1991, 2001; Poulin-Dubois et al., 1996; Poulin-Dobois, 1999). Meltzoff (1995) also claimed that infants restrict their mental state attributions to people. In his study, when a human agent was replaced by a set of mechanical pincers, children failed to reproduce the incomplete action (Meltzoff, 1995: Experiment 2). However, Johnson, Booth, and O'Hearn (2001) replicated this experiment using a stuffed orangutan, demonstrating that a nonhuman agent could elicit the re-enactment of goal-orientated behavior by an infant under certain circumstances. The authors concluded that the agent needed to possess the features thought to characterize mentalistic agents, such as the ability to interact contingently with others, or the presence of a face (Johnson, Booth, and O'Hearn, 2001).

In the present study, when young children saw the robot "try" but fail to achieve the same set of target outcomes, and were given the objects they had seen the robot manipulate, they produced the inferred outcome, rather than the actually viewed event only when the robot showed intention-like actions, such as gazing. Following Meltzoff (1995), these responses were interpreted as evidence that the infant attributed goals to the agent. However, the children who saw the robot that did not gaze at its partner or the object in the same action condition (the incomplete condition) did not produce the target outcomes. This contrasts with the results of Meltzoff's (1995) study, in which the human demonstrator was not required to exhibit behavior implying an intention, such as gazing at the objects, in order to induce infants to produce the target outcome. These variances could be based on the ability of children to distinguish between humans and nonhumans.

Experiment 2: False belief task with a robot by preschoolers

Method

Participants: The participants were 58 young children (27 boys, 31 girls; range=54 months to 80 months; mean=65.4 months). We chose children of these ages because many studies demonstrated that children between the ages of four and five years start to pass the False Belief Task.

Materials: All the stimuli were presented on a video monitor. There were two versions of video stimuli. One of the video scenes was as follows: Robovie (see Experiment 1) puts the doll away in a particular location (Box A), then leaves the room. During Robovie's absence, the man removes the doll from Box A, and places the doll not back in Box A, but in Box B. The other video scene was the same as the robot version, except that a human was projected, instead of the robot.

Stimuli: Each subject was shown these two types of scenes, and given four questions just after watching each video scene individually. The order of presenting the stimulus was counterbalanced. Four questions are as follows: i) "Where will it/he look for a doll?" (Question for prediction); ii) "Which box does it/he think the doll is in?" (Question for representation); iii) "Which box contains a doll?" (Question for reality); iv) Which box contained a doll at the beginning of the session?" (Question for memory).

Results and discussion

The results are shown in Fig. 2. There was no difference between the human condition and the robot condition in the reality question ($z=0.01$, $P>0.992$, n. s.) and the memory question ($z=0.28$, p>o.339, n. s.). Most of the children answered these questions correctly. There was also no difference in the prediction question ($z=0.28$, $p>0.339$, n. s.) between both conditions. However, there was a significant difference between the human condition and the robot condition in the representation question ($z=3.68$, $p<0.003$). These results show that by the ages of four to five children attribute false belief to a robot but they do not attribute a mental verb to it. This means that, for children, the robot does not have thinking capabilities or thoughts in this situation.

In conclusion, we demonstrated that young children discriminate between a robot and a human in mentalizing when the mental verb was used in a question such as "think" in a False Belief Task. It seems to be difficult for young children to link the behavior of just searching and thinking in a robot.

Figures

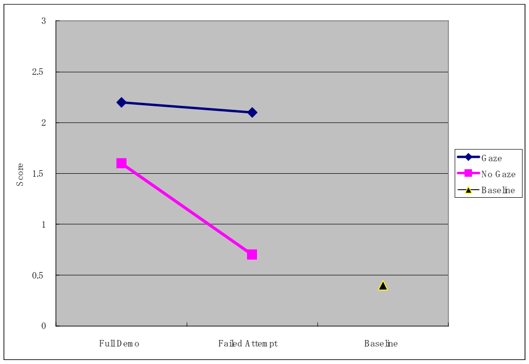

Figure 1: Results of Experiment 1.

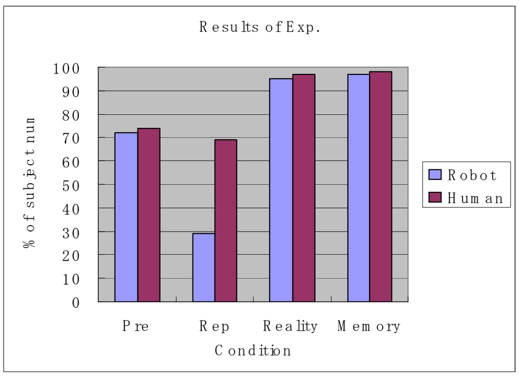

Figure 2: Results of Experiment 2.

Acknowledgments

These studies were supported by the ATR Intelligent Robotics and Communication Laboratory, and by a grant from JSPS (No: 13610087, 16500161) to Shoji Itakura.

References

Barr, R. & Hayne, H. (2000). Age-related changes in imitation: implications for memory development. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne, (Eds.), *Progress in infancy research*. Hillsdale, NJ: Lawrence Earlbaum Associates.

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).

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 447–458.

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 459–473.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Science, 4*, 22–28.

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

Johnson, S. C., Booth, A., & O'Hearn, K. (2001). Inferring the goals of non-human agents. *Cognitive Development, 16*, 637–656.

Legerstee, M. (2001). Domain specificity and the epistemic triangle: The development of the concept of animacy in infancy. In F. Lacerda, C. von Hofsten, & M. Heimann (Eds.), *Emerging cognitive ability in early infancy*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *Journal of Experimental Child Psychology, 51*, 423–433.

Meltzoff, A. N. (1995). Understanding the intention of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.

Phillipes, A. T. & Wellman, H. M. (in press). Infants' understanding of object-directed action. Cognition.

Poulin-Dubois, D. (1999). Infants' distinction between animate and inanimate objects: the origin of naïve psychology. In P. Rochat (Ed.), *Early social cognition.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development, 11*, 1-16.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.