Running Head: word learning from a robot

Can young children learn words from a robot?

Yusuke Moriguchi[a], Takayuki Kanda[b], Hiroshi Ishiguro[c], Yoko Shimada[d] & Shoji Itakura[e]

[a] Department of School Education, Joetsu University of Education

[b] ATR Intelligent Robotics and Communication Laboratories

[c] Department of Adaptive Machine Systems, Osaka University

[d] Department of African Area Studies, Kyoto University

[e] Department of Psychology, Graduate School of Letters, Kyoto University

Correspondence should be addressed to Yusuke Moriguchi, 1 Yamayashiki-machi, Joetsu 943-8512, Japan. E-mail: moriguchi@juen.ac.jp

Published in *Interaction Studies*

Abstract

Young children generally learn words from other people. Recent research has shown that children can learn new actions and skills from nonhuman agents. This study examines whether young children could learn words from a robot. Preschool children were shown a video in which either a woman (human condition) or a mechanical robot (robot condition) labeled novel objects. Then the children were asked to select the objects according to the names used in the video. The results revealed that children in the human condition were more likely to select the correct objects than those in the robot condition. Nevertheless, the five-year-old children in the robot condition performed significantly better than chance level, while the four-year olds did not. Thus there is a developmental difference in children's potential to learn words from a robot. The results contribute to our understanding of how children interact with non-human agents.

developmental cybernetics, word learning, social cognition, cognitive development

Introduction

Today robots substitute for over a million workers, primarily in industrial or work settings. In the future, robots may perform household chores, and perhaps even play a role in child care and education. But we currently know little about how children interact with robots. For example, we do not know whether young children can effectively learn from a robot.

The emerging research field of *developmental cybernetics* examines interactions between children and robots, and builds a theoretical framework regarding the characteristics that facilitate these interactions, in areas such as learning and teaching (Itakura, Okanda, & Moriguchi, 2008; Kojima, 2005). For example, research has shown that young children prefer more interactive robots to non-contingent robots (Arita, Hiraki, Kanda & Ishiguro, 2005; Moriguchi, Kanda, Ishiguro & Itakura, 2010; Tanaka, Cicourel & Movellan, 2007). Research has given particular attention to young children's learning from robots, such as whether and how children learn actions and skills from a robot as well as they do from a human (Itakura, Ishida et al., 2008: Johnson, Booth, & O'Hearn, K. 2001; Moriguchi, Kanda, et al., 2010). For example, Itakura, Ishida et al. addressed this issue using an imitation paradigm developed by Meltzoff (1995). In this study, children watched videos showing a humanoid robot that manipulates an object (e.g., pulling a dumbbell apart). The robot performed the actions successfully in Successful Demonstration trials, and failed to achieve the goal in Unsuccessful Demonstration trials (e.g., the dumbbell remained intact). In one condition (Eye Contact condition), the robot made eye contact with the experimenter on video before and after it failed the action. In another condition (No Eye Contact condition), the robot did not make eye contact with the experimenter. The results revealed that children in both conditions successfully reproduced the observed actions in the Successful Demonstration trials, but only children in the Eye Contact condition reproduced the robot's "intentional" actions in the Unsuccessful Demonstration trials. These studies suggest that children could learn actions from a robot, and that social signals, such as eye contact, may facilitate children's imitative learning, which is consistent with other findings that ostensive cues may be important for infants' learning from another person (Csibra & Gergely, 2009).

Despite the accumulating evidence of action imitation in developmental cybernetics, another type of learning, word learning, has largely been unstudied. Recent articles on word learning underscore the importance of social interaction, where the presence and actions of a speaker provide context and play an important role in childhood language development

(Baldwin, 2000; Tomasello & Barton, 1994). Specifically, there is growing evidence that infants and young children rely on a speaker's social signals such as gaze and head movement, and mental states such as knowledge states (Baldwin, 1991; Baldwin, & Moses, 2001; Sabbagh & Baldwin, 2001; Tomasello & Barton, 1994). However, it has been also shown that children may build their vocabulary not only from human interactions but also from multimedia sources, such as television or video (Anderson & Pempek, 2005; Krcmar, Grela, & Lin, 2007; Rice, Huston, Truglio, & Wright 1990; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009; Wright et al., 2001). A recent study showed that older preschool children can learn verbs from a video (Roseberry et al., 2009), but it also showed that live interaction with an experimenter facilitated children's word learning from a video. These studies indicate that social interaction with another person is important, but not always necessary, for children's word learning. The present study builds upon this previous research and examines whether children are able to learn words from robots.

Only one study reported about children successful learning words from a robot. Kanda, Hirano, Eaton, and Ishiguro (2004) examined whether Japanese school-aged children could form relationships with robots and whether the interactions with robots facilitated children's second language acquisition (English). In this study, children interacted with English speaking robots for two weeks in their classroom. Children were given an English test before and after the interactions with the robots. The results revealed that the children who spent more time interacting with the robots scored better on the English posttest. The present study extends these findings in two ways. First, the previous study was conducted as a field trial and showed improvement on English tests. However it is unclear whether the improvement was actually due to the interactions with the robots. A high-interactive child may have also interacted more with other children and learned words from the children. A more controlled experiment is needed to assess whether the children actually learned the words from a robot. Therefore, we conducted a basic word learning experiment where children observed a robot labeling novel objects with novel words. Second, the previous study did not address whether the children learn words from a robot as well as they do from a human. Moriguchi, Kanda et al. (2010) showed that children did not learn actions from a robot as readily as they did from a human. Therefore, we examined whether the efficiency in word learning differs between a robot and a human.

O'Connell, Poulin-Dubois, Demke, and Guay (2009) addressed the issues, examining whether infants can establish word-object relations using a robot’s gaze direction. In the study, infants were exposed to a robot who gave a novel label for a novel object. As the result, infants followed the robot's gaze toward the word referent, but failed to establish word-object relations. The study suggested that infants may have difficulty with word learning from a robot. Given the evidence, the present study selected four- and five-year-old children as participants. Children watched either a woman or a robot labeling three novel objects separately in Japanese (e.g., "This [object] is a toma."). Then, the children were presented with the objects and asked to point to the correct object in response to the evaluator's instructions ("Which [object] is a toma?"). An autonomous robot was used in this experiment (see Figure 1). The woman's voice was recorded and played by the robot. This method allowed us to minimize the difference between the human and the robot. Before the test phase, a control phase was inserted. During the control phase, the agents introduced themselves to the children, stating their name and favorite food. Then, the children were asked control questions about the agents' name and favorite food. The questions were intended to address whether the children were able to interpret information presented by the agents.

Method

Ethics Statement

Participants were recruited from nursery schools in Fukuoka. Informed consent was obtained from the parents and teachers of children prior to their involvement in the study. This research was conducted under strict ethical guidelines and the study design was approved by the ethics review board at the Department of Psychology, Kyoto University.

Participants

Thirty-four four-year-old children (between 50 and 61 months of age, $M = 54.9 \pm 3.6$ months; 12 girls), and 30 five-year-old children (between 63 and 74 months of age, $M = 68.3 \pm 3.6$ months; 10 girls) participated in this study. The children were randomly assigned to one of two conditions: a human condition and a robot condition. There were no significant age differences among participants assigned to the two conditions.

Stimuli

A robot named Robovie was used as a speaker. Robovie was developed at the ATR Intelligence Robotics Laboratory in Kyoto, Japan. It is an autonomous humanoid robot (1.2 m in height, 50 cm in diameter, and 40 kg in weight) with human-like eyes and hands (see Figure 1). The same robot was used in other studies (Moriguchi, Kanda et al., 2010).

Video clips were created for the control phase and the test phase. In the control phase, a woman (human condition) or robot (robot condition) began with an introduction, saying their name and favorite food (banana). In the test phase, the agents were presented with a novel object, and they labeled the object with a novel word (toma, blicket, or hano). During the test phases, both agents moved its head and gazed at the object, and then labeled the objects. There were three short video clips for the test phase, where each novel object was labeled with a novel word. The voice of the agent (recorded woman's voice) and the objects and their assigned names remained the same in both conditions.

Procedure

Each child was tested individually for approximately 15 minutes. The child was seated at a table, and the evaluator sat at a table next to the child. The evaluator spoke briefly with the child to establish rapport. Once the child appeared relaxed, the experiment began. This task had two phases: the control phase and the test phase.

At the beginning of the control phase, the child was instructed that s/he was going to learn some words from agents: "Now, we are going to learn some objects' names from a lady (or robot) on the video. Before that, the lady (or the robot) will introduce themselves. Please watch the video carefully. Okay?" Then, the children watched the video clip twice. In the video, the agent said: "My name is (name). I like bananas!" After watching the video, the child was given the control questions: "What is her (or the robot's) name? What is her (its) favorite food?" The control questions were used to ascertain whether the child focused on the words spoken during the video and whether s/he was able to interpret that information. The child was regarded as having passed the control phase when s/he answered the "favorite food" question correctly.

In the test phase, the child was presented a novel object and asked to name the object: "Do you know the name of the object?" Initially, none of the children were able to answer this question with the appropriate name. When the child answered "No," s/he was instructed to watch the new video clips: "Now, she (the robot) is going to label the object. Please watch the video and listen to her (its) voice carefully." During the clip, the agent was presented with a novel object and the agent labeled it with a novel word while gazing at the object, "This is a (label)". The child watched each video clip twice. After viewing the video, the child was presented with a new object, and watched the next video clip. The child was given three trials consecutively. The order of the objects presented was counterbalanced. Following a short break, the evaluator then presented the child with the three objects labeled in the video clips and gave him/her test questions (e.g., "Which is a toma?"). The child was asked to point to the correct object in response to the evaluator's questions. The three test questions were given consecutively.

Results

We first analyzed the results for the control questions. Children were scored based on correctly answering the control question, specifically if they could name the agent's favorite food (bananna). Almost all the children in both the human and robot conditions answered the control question quite easily. Indeed, 94% of the four-year-old children and all of the five-year-old children in the study correctly interpreted the information from both agents (Figure 2A). One child from each condition failed to answer the control question correctly and was subsequently excluded from further analyses.

Next, we analyzed the results for the test questions. Children were scored 1 when they correctly responded to the evaluator's instructions by pointing to the target object (total score range 0-3). The children in the human condition performed quite well (Figure 2B). The four-year-old children in the human condition pointed to the correct object for an average of two out of the three questions ($M = 2.13$), and the five-year-old children scored almost perfectly ($M = 2.73$). On the other hand, the children in the robot condition showed some difficulty with the task. The mean score for the four-year-old children in the robot condition was 1.25, while the mean score for the five-year-old children was 2.33. The scores on the test questions were analyzed using an age (four-year-old vs. five-year-old) ×

condition (human vs. robot) two-way ANOVA. There were significant main effects of age, F (1, 58) = 9.309, $p < .003$, ή = .14, and condition, F (1, 58) = 5.288, $p < .025$, ή = .08. Five-year-old children were more likely than four-year-old children to perform the task correctly. The children in the human condition performed significantly better than those in the robot condition. The interaction between age and condition was not significant, F (1, 58) = 0.74, $p > .39$, ή = .01.

We also examined whether the performance of children in each group was significantly different from what would have been expected to occur by chance. The chance level for each trial was 33%; therefore, the score expected by chance was 1. A one-sample *t*-test revealed that the four-year-old children in the robot condition performed at chance level on the test questions, t (15) = 0.845, $p > .10$, whereas the four-year-old children in the human condition correctly identified the objects above chance level, t (15) = 3.737, $p < .002$. The five-year-old children in both the human condition and the robot condition selected the correct objects at greater than chance levels, t (14) = 9.539, $p < .001$ and t (14) = 4.394, $p < .001$, respectively.

Finally, we examined individual response patterns in each group. Children were classified into a word-learner according to whether children performed all three trials correctly in the test phases. As the results, ten out of sixteen 4-year-old children in the human condition (62.5%) and four out of sixteen children in the robot condition (25 %) were regarded as the word-learner. We conducted a chi-square test to examine the differences in the frequencies between conditions, and found the significant differences between conditions, χ^2 (1, N = 32) = 4.571, $p < .05$. On the other hand, most of the 5-year-old children were classified into the word-learner (13 out of 15 in the human condition and 11 out of 15 in the robot condition). We found no significant differences between conditions, χ^2 (1, N = 30) = 0.833, $p > .36$.

Discussion

The results of this study showed that the five-year-old children in the robot condition successfully performed in the test phase significantly above chance level. Although their performance was poorer than the five-year-old children in the human condition, the children associated the voice from the robot with the objects they were identifying. This is evidence for children learning words from a robot. In this study, children learned words not only from a human but also from a robot. The results suggest that a nonhuman agent may potentially play an important role in vocabulary development.

The present study also showed that young children were more likely to learn words from a woman than a mechanical robot, even though the acoustic cues (i.e., speech) were the same for both agents. Children in the human condition associated the model's labels with the novel objects more easily, resulting in higher performance when prompted to select the named objects. On the other hand, children in the robot condition found it relatively more difficult to select the correctly named objects during the test phase. The results suggest that an agent's appearance may be a factor for the development of language in young children. Children may learn words more efficiently from a robot with more human-like appearance (e.g., android). Indeed, a previous study showed that children may learn actions from an android more efficiently than from a robot. (Moriguchi, Minato, Ishiguro, Shinohara, & Itakura, 2010).

A question remains as to why the children in the robot condition performed worse than those in the human condition, especially the four-year-old children. In other words, why did the difference in appearance between a human and a robot produce different performance among children in the test phase? There are two possible interpretations. One simple explanation is that children have less experience with a robot compared to a human, and this lack of experience may induce additional effort (or working memory demand) for a child. This may also explain why the children in the robot condition improved their performance between four and five years of age. Extensive research has shown that children develop executive function, such as working memory and inhibition control, during their preschool years (Carlson, Moses, & Breton, 2002; Garon, Bryson, & Smith, 2008; Moriguchi, Lee, & Itakura, 2007; Zelazo, Müller, Frye, & Marcovitch, 2003). Thus the development of their working memory may enable children in the robot condition to improve their word learning between ages four and five. It could be that reducing the number of words (i.e., reducing the task demand) should improve the 4-year-olds' performances in the robot condition.

The second possible explanation is that a human’s actions, unlike a robot’s actions, induce young children’s imitative process (or learning process), which can affect children’s word learning. It has been shown that observing human actions, but not mechanical actions, may elicit young children’s and adults’ imitative behavior (Kilner, Paulignan, & Blakemore, 2003; Meltzoff, 1995; Tai et al., 2004). Some researchers suggest that young children may mentally simulate a human’s actions while observing the actions, and therefore could reproduce the human’s actions more easily after observing such actions (Moriguchi, Kanda et al., 2010). This may also be true for the present study. Children may have some difficulty with mentally simulating the labeling while observing a robot labeling an object, which may have caused the children's poorer performance during the test phase. However, it is still unclear what cognitive processes are involved in word learning that produce the differences between learning from a human compared to from a robot. These questions should be addressed in future research.

One might argue that the results of the human and robot conditions may vary due to the fact that children in the robot condition may be distracted by the robot's actions because the robot is novel and attracts attention. However, the children in the robot condition performed well in the control phase, equally well as the children in the human condition. Also, another study showed that when presented with a robot movie for the first time (the same robot as in the present study), preschool children paid close attention to the robot's actions (Moriguchi, Kanda et al., 2010). Children in this study were instructed to observe a robot sorting cards according to one particular rule (e.g., color rule). After the observation, the children were then asked to sort the cards according to the same rule. Importantly, children were not given instructions about the specific rule used by the robot. In this study, preschool children reproduced the robot's actions almost perfectly. On the basis of this evidence, children in the robot condition were found to be paying close attention to the actions of the robot.

Conclusion

The present study extends previous findings of word learning in developmental cybernetics in important ways. A previous study showed in a field trial that children can improve their performance on a test of a second language through interactions with a robot (Kanda et al., 2004). We used a better controlled experiment and showed that children can learn words from a robot in their first language. Also, we found that children did not learn words from a robot as efficiently as they did from a human. However, there are several questions that should be addressed. First, it remains unclear whether children younger than five-years-old can learn words from a robot. Our study found that four-year-old children failed to learn words from the robot. Given that children aged 18 months failed to learn words from a non-human agent (O'Connell et al., 2009), it seems likely that children may begin to learn words from a robot during preschool ages. Further research should be necessary to address whether younger children could learn words from a robot. Second, understanding how the appearance of a robot affects children's ability to learn should be addressed. In this study, we used a mechanical robot with human eyes and hands as a speaker for the child. The robot was different from the human condition in many ways, such as face and body. Thus, further study is needed about which aspects of appearance have the strongest impact on children's learning. Third, the present study addressed how children associated novel labels with novel objects, but did not address whether children learn words from a human or a robot in a broader sense. For example, it is unclear whether children retained the labels after a long period. These points can be addressed using longitudinal method. Finally, it may be of interest to examine individual differences in word learning among children using the same robot. Some children learned words from the robot, while others did not. Children's socio-cognitive abilities, or working memory, may be linked to the individual differences that were found. Future research should focus on individual differences in children's learning abilities by covarying verbal skills and word learning.

Acknowledgements

The first author received support from a JSPS Research Fellowship for Young Scientists. This research was also supported by grants from the Nissan Science Foundation to Shoji Itakura. The authors would also like to thank the children who participated in the study.

References

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist, 48*, 505 – 522.

Arita, A. Hiraki, K., Kanda, T., & Ishiguro, H. (2005). Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition, 95*, B45- B57.

Baldwin, D.A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875-890.

Baldwin, D.A. (2000). Interpersonal understanding fuels knowledge acquisition. *Current Directions in Psychological Science, 9*, 40-45.

Baldwin, D.A., & Moses, L.M. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309–329.

Carlson, S.M., Moses, L.J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73-92.

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

Garon, N., Bryson, S.E., & Smith, I.M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

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

Itakura, S., Okanda, M., & Moriguchi, Y. (2008) Discovering mind: Development of mentalizing in human children. S. Itakura & K. Fujita (Eds.), *Origins of social mind: Evolutionary and developmental view* (pp.179-198). Springer.

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

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004) Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human Computer Interaction, 19*, 61-84.

Kilner, J.M., Paulignan, Y., & Blakemore, S.J. (2003). An interference effect of observed biological movement on action. *Current Biology, 13*, 522-525.

Kojima, H. (2005). *Children-robot interaction: From interaction to cognition.* Paper presented at the ESF research conference on brain development and cognition in human infants “From action to cognition". Acquafredda di Maratea, Italy.

Krcmar, M., Grela, B. G., & Lin, Y. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*, *10*, 41–63

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

Moriguchi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2010) Children perseverate to a human’s actions, but not to a robot’s actions. *Developmental Science*, *13*, 62-68.

Moriguchi Y., Lee, K., & Itakura S. (2007). Social transmission of disinhibition in young children. *Developmental Science*, *10*, 481- 491.

Moriguchi, Y., Minato, T., Ishiguro, H., Shinohara, I. & Itakura, S. (2010). Cues that trigger social transmission of disinhibition in young children. *Journal of Experimental Child Psychology, 107,* 181-187.

O'Connell, L., Poulin-Dubois, D., Demke, T., & Guay, A. (2009) Can infants use a nonhuman agent’s gaze direction to establish word-object relations? *Infancy 14*, 414-438.

Rice, M.L., Huston, A.C., Truglio, R., & Wright, J. 1990. Words from "Sesame Street": Learning vocabulary while viewing. *Developmental Psychology, 26*, 421-428.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development, 80*, 1360-1375.

Sabbagh, M.A., & Baldwin, D.A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.

Tai, Y.F., Scherfler, C., Brooks, D. J., Sawamoto, N. & Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Current Biology, 14*, 117–120.

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

Tomasello, M., & Barton, M.E. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*, 639-650.

Wright, J.C., Huston, A.C., Murphy, K.C., St. Peters, M., Pinon, M., Scantlin, R., & Kotler, J. 2001. The relations of early television viewing to school readiness and vocabulary of children from low-income families: The Early Windows Project. *Child Development, 72*, 1347-1366.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68 (3),* Serial No. 274.

Figure Legends

Figure 1. The robot used in the study

Figure 2. Results. *(A)* Pass rate in the control phase, and *(B)* mean correct answers given in the test phase