Easy or Not Easy: Young Children’s False Belief Understanding in Communicative Situations

Kensuke Sato (kensuke@p.u-tokyo.ac.jp)
Department of Educational Psychology, Graduate School of Education
The University of Tokyo, Japan

Abstract

The present study investigated whether young children are better at representing others’ false beliefs in communicative situations than in standard false belief tasks. Younger (3.7–4.6 year-olds, n = 18), middle-aged (4.7–5.6 year-olds, n = 20), and older children (5.7–6.6 year-olds, n = 20) completed both standard false belief and communicative tasks having the same story structure. Although the children gave fewer correct answers in the communicative tasks, they expressed more uncertainty toward their choices in the communicative tasks than in the standard tasks. The results appear to indicate that children belonging to the younger group are more sensitive to others’ knowledge states in communicative settings than in hypothetical situations. The results also demonstrate the difficulty in inhibiting the literal meaning of utterances for children.

Keywords: theory of mind; young children; false belief; communication; literal meaning

Introduction

Understanding others’ mental states essentially refers to our ability to communicate our intentions, desires, and beliefs to one another. When and how this ability is acquired has been the focus of a number of investigations. In these developmental studies of mind-reading ability (or “theory of mind”), gaining an understanding of others’ false beliefs is considered to be a crucial turning point because it provides strong evidence for the appreciation of the distinction between mind and reality. Thus, many researchers have focused on children’s understanding of false beliefs by using false belief tasks. For example, a well-known false belief task, the Sally-Ann task (Baron-Cohen, Leslie, and Frith, 1985), involves presenting the following scenario to participants. A doll named Sally puts her ball in a box and leaves the room. While Sally is away, a doll named Ann moves the ball from the box to a basket. Then, Sally returns to the room. After the participant has seen this scenario being enacted, he/she is asked where Sally will look for her ball: in the box, or in the basket. Four-year-olds and older children often judge that Sally will search for the ball in the box, although the ball is actually in the basket. These actions prove that the children know that Sally’s actions depend on her beliefs rather than on what actually took place, which the children are aware of. On the other hand, many 3-year-olds typically fail such tasks and judge that Sally will search for the ball in the basket. Wellman, Cross, and Watson (2001) conducted a meta-analysis of more than 100 false belief studies including various kinds of false belief tasks and found that children aged 3.5 years and less consistently commit the realist error as if they are unable to distinguish the belief of the character of a story from reality.

In contrast to many false belief task studies, 3-year-olds and younger children are highly sensitive to others’ minds in naturalistic situations (e.g., O’Neill, 1996; Bartsch & Wellman, 1995; Dunn, 1988). For example, in their natural interaction, even 2-year-olds talk and behave as if they understand the beliefs of others (Dunn, 1988). This discrepancy between children’s false belief task performance and the nature of their communication implies that children’s mind-reading ability is masked in artificial experimental situations. This observation is now evidenced by more direct investigations focusing on children’s false belief understanding in communicative situations. Communicative false belief tasks, which investigate children’s understanding of the false belief of the person with whom they communicate, are much easier than standard tasks constructed with a hypothetical story (Bezuidenhout & Sroda, 1998; Carpenter, Call, & Tomasello, 2002; Happe & Loth, 2002). For example, Carpenter et al. (2002) conducted a study involving communicative tasks such as the following. In full view of the participant and one experimenter (E1), another experimenter (E2) places object X in a container and object Y beside the container. Then, E1 leaves the room. In E1’s absence, E2 takes object X out of the container and places object Y in that container. Next, E2 takes Y out of the container, and places X and Y on the blanket in the room. Then, E1 returns. E1 tries to open the container but fails and finds the objects on the blanket. E1 asks the participant to bring the “toy” he/she wanted to him/her. Surprisingly, very young participants aged 2.8–3.2 years took the object X that E1 had falsely believed was in the container in about 70% of the trials. Therefore, Carpenter et al. (2002) claimed that communication leads children to attain the mindset of mind-reading. In other words, they support the hypothesis that young children are better at representing others’ false beliefs in communicative settings than during standard tasks.

However, the hypothesis that communication helps young children’s false belief understanding is not supported by sufficient evidence. First, this hypothesis has never been tested by means of a fair comparison. Previous studies have only involved communicative tasks or compared children’s performance in communicative tasks and standard tasks that had completely different story structures, that had a different number of story characters and objects, and that even had different probabilities of the children obtaining the correct response by chance. Therefore, a fair comparison—or more specifically, a comparison between standard tasks and communicative tasks that only differ on the basis of whether or not they involve “communication”—is necessary.

Second, due to the insufficient number of studies, the extent to which this hypothesis (i.e., that communication helps young children’s false belief understanding) can be
generalized is unknown. For example, in verbal communication, the listener sometimes has to inhibit the literal meaning of an utterance. Consider a person who falsely believes that object A is in Place 1 and object B is in Place 2, when the places of the objects are actually the reverse. The person states, “The object is in Place 1.” In order to interpret this utterance correctly, the listener, who knows the actual location of the objects, needs to inhibit the literal meaning of “The object is in Place 1” and interpret this utterance as a reference to the object in Place 2. Perhaps, due to underdeveloped central executive functions, young children sometimes display difficulty in the inhibition of literal meaning (e.g., Epley, Morewedge, & Keysar, 2004; Noveck, 2001). The abovementioned procedures of the communicative tasks by Carpenter et al. (2002) did not impose the inhibition of literal meaning on the participants because they could not interpret the critical utterance literally (the participants could not specify the referent of the word “toy” in E1’s critical utterance, from their egocentric perspectives). Even the communicative tasks used in the other studies did not impose the inhibition of literal meaning on participants. Therefore, the hypothesis by Carpenter et al. (2002) has never been tested in a situation imposing the inhibition of literal meaning on children. The effect of the inhibitory process in communicative false belief tasks has not been investigated in previous studies. This study aims to address the abovementioned problems. The first question is as follows: When the standard task and communicative task have the same story structure, is the communicative task easier for young children? To test—in a fairer manner—the hypothesis that communicative situations help young children to appreciate others’ false beliefs, I used standard and communicative false belief tasks with the same story structure. Due to the fact that the participants perform both the tasks, this study tests the influence of the factor of communication more directly. The second question is as follows: When the communicative task imposes the inhibition of literal meaning on children, do young children find this task easier than the standard task? This study also investigates the effect of the inhibition of literal meaning on performance in a communicative task. For this purpose, I developed two types of communicative tasks: literal and nonliteral.

Method

2.1 Participants

The participants were 60 young children. They were divided into three age groups, each group having 20 children: 3.7–4.6 years, 4.7–5.6 years, and 5.7–6.6 years (referred to as the “younger,” “middle-aged,” and “older” groups, respectively). All the participants were monolingual Japanese students of a preschool in an urban district of Kanagawa. The data collected from 58 participants (18 children of the younger group) who passed the memory control questions were used for the analysis.

2.2 Material

Four objects that were unfamiliar to the participants and four containers were used in two communicative tasks: Two objects and two containers were used for the literal task, and the remaining two objects and two containers were used for the nonliteral task. The two pairs of containers differed only in terms of color. A 13-inch MacBook display and PowerPoint were used to conduct the two standard tasks.

2.3 Procedure

All the participants individually completed two standard tasks (the Sally-Ann task and matched standard task) and two communicative tasks (the literal matched communicative task and the nonliteral matched communicative task). The order of the tasks was counterbalanced for each age group. The action-desire discrepant task (ADD), originally created by Robinson and Mitchell (1994) (and also used in Mitchell, Robinson, and Thompson, 1999), was used as the matched standard task. Two communicative tasks with the same story structure as the ADD were developed. The story of the three matched tasks was as follows. P1 and P2 are hypothetical story characters (matched standard task) or experimenters (two communicative tasks).

In full view of the participant and P1, P2 places one object in container A and the other object in container B. Then, P1 leaves the room. In P1’s absence, P2 switches the locations of the objects. Then, P1 reenters the room.

In the two communicative tasks, two experimenters played P1 and P2, using real objects and containers. In the matched standard task, the story was presented in animated form, using PowerPoint. The three matched tasks had different endings and critical test questions. At the end of the matched standard task, the participants saw that P1 approached one of the containers and were asked to select the object—out of the two pictures depicting the objects—that P1 wanted. In the two communicative tasks, P1 called the children to her at the entrance of the room. When a child came to her, P1 verbally requested the child to bring her one of the objects by referring to the container (e.g., “Hey, XX (participant’s name)! The toy in the red box is mine! Please bring me my toy in the red box.”). Thus, all the three matched tasks had test questions, which imposed (on the children) the reasoning about P1’s desired objects, based on her false belief and actions (physical actions in the matched standard task and speech acts in the two matched communicative tasks). The participants who chose the object that had been in the container that was indicated or that had been approached by P1 before she left the room were coded as correct in each task.

The two communicative tasks, literal and nonliteral, differed with regard to the following point. In the nonliteral task, while P1 spoke to the participant at the entrance, P2 silently removed the objects and placed them in front of each container. That is, when the participants tried to take the object, there was no object in the containers.
Thus, they could not take the object in the container. In this manner, these characteristics of the task were considered to make it impossible for the participants to behave literally (e.g., take the toy in the red box), and as a result, were considered to reduce the processing effort involved in inhibiting the literal meaning of the critical utterance of P1. P1 and P2 never named the objects in any of the matched tasks; instead, the objects were indicated using the pronouns “this” or “that” or were simply referred to as “toy.” For the Sally-Ann task, I used the same story as that used in the original task of Baron-Cohen, Leslie, and Frith (1985). The location of the correct object (right or left) was also counterbalanced for each age group. All the sessions were videotaped. Table 1 presents a comparison of the task types and story types in the four different tasks used in this experiment.

### Results

#### 3.1 Analysis of the proportion of correct responses

Table 2 shows the proportions of correct responses of the children in the four types of false belief tasks. As many studies have reported, the proportion of correct responses in the two standard tasks revealed strong age trends. Despite the somewhat aberrant way of story presentation in the Sally-Ann task and matched standard task (the tasks were presented in animated form on the MacBook display), the results revealed the same pattern as those obtained in many other false belief studies. The children found the matched standard task slightly more difficult than the Sally-Ann task because of the additional object and slightly complicated story structure. In the two communicative tasks, however, the proportion of correct responses did not increase with age. These observations were also supported by the statistical analysis. Chi-square tests were conducted for between-group comparisons in each task and the results revealed that the number of correct responses significantly differed across the age groups in the two standard tasks (Sally-Ann: $\chi^2 (2) = 30.67, p < .01$; matched: $\chi^2 (2) = 12.96, p < .01$). Ad hoc multiple comparisons revealed that the older and middle-aged groups scored higher than the younger group in the Sally-Ann task, and in the matched standard task, the older groups performed better than the middle-aged and younger groups (Ryan’s procedure: $ps < .05$). The number of correct responses in the two communicative tasks did not differ across the age groups (literal: $\chi^2 (2) = 1.80, p = .41$; matched: $\chi^2 (2) = 0.90, p = .64$).

The main purpose of this study was to compare the children’s performances in the standard task and the communicative tasks, both having the same story structure. Between-task comparisons in each age group revealed a significant difference in the performances of the older and middle-aged groups (Cochran’s Q test for the older group: $Q (3) = 3.40, p < .05$; middle-aged group: $Q (3) = 3.40, p < .05$).

Multiple comparisons revealed that among the older group, more correct answers were obtained in the two standard tasks than in the two communicative tasks. The middle-aged group provided more correct answers in the Sally-Ann task than in the two communicative tasks (Ryan’s procedure: $ps < .05$), while no significant difference in the number of correct responses was observed in the younger group ($Q (3) = 3.40, p = .33$). In this analysis, the hypothesis that communicative situations help young children’s mind-reading was not supported. The communicative setting did not increase the proportion of correct responses in any age group; in fact, it lowered the proportion of correct responses in the older group. This study also aims to address the effect of “literalness” in the communicative false belief tasks; however, this effect was not confirmed in this analysis. No significant difference was observed between the performances in the matched literal and matched nonliteral tasks.

### Table 1: Summary of the differences in task type and story type across the four types of false belief tasks

<table>
<thead>
<tr>
<th>Task type</th>
<th>Story type</th>
<th>Sally-Ann</th>
<th>Matched standard</th>
<th>Matched literal</th>
<th>Matched nonliteral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Presented on the MacBook display</td>
<td>22.2%</td>
<td>22.2%</td>
<td>38.9%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Communicative</td>
<td>Communication with two experimenters</td>
<td>Middle-aged (n = 20)</td>
<td>85.0%</td>
<td>45.0%</td>
<td>35.0%</td>
</tr>
<tr>
<td>Communicative</td>
<td>With two objects in two containers respectively</td>
<td>Older (n = 20)</td>
<td>100%</td>
<td>80.0%</td>
<td>20.0%</td>
</tr>
</tbody>
</table>

#### Table 2: Proportions of correct responses of the children from each age group in the four types of false belief tasks

<table>
<thead>
<tr>
<th>Group</th>
<th>Sally-Ann</th>
<th>Matched standard</th>
<th>Matched literal</th>
<th>Matched nonliteral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger (n = 18)</td>
<td>22.2%</td>
<td>22.2%</td>
<td>38.9%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Middle-aged (n = 20)</td>
<td>85.0%</td>
<td>45.0%</td>
<td>35.0%</td>
<td>40.0%</td>
</tr>
<tr>
<td>Older (n = 20)</td>
<td>100%</td>
<td>80.0%</td>
<td>20.0%</td>
<td>30.0%</td>
</tr>
</tbody>
</table>
This study also aims to address the effect of “literalness” in the communicative false belief tasks; however, this effect was not confirmed in this analysis. No significant difference was observed between the performances in the matched literal and matched nonliteral tasks.

3.2 Analysis of the expressions of uncertainty

The analysis of the proportion of correct responses does not support the hypothesis; in fact, it suggests the reverse. At this point, I conducted another analysis of the children’s expressions of uncertainty. Generally, not only the correct response but also the children’s expressions of uncertainty are considered as a meaningful measure of sensitivity to the other’s false belief (Carpenter et al., 2002). If a child cannot distinguish the other’s belief from reality, he/she has committed a realist error without hesitation (e.g., in the communicative tasks, the child immediately took the object in the container that was referred to literally). Therefore, children’s expressions of uncertainty were also employed as a measure of children’s false belief understanding. To determine whether children were uncertain about their choice, a checklist comprising behaviors evidencing uncertainty was prepared for each task. The behaviors included in the checklist were (a) looking questioningly at the experimenter, (b) verbally asking for the experimenter’s help, (c) clarifying the meaning of the test question, (d) changing one’s first choice, (e) trying to choose both alternatives, (f) hesitating for a long time with regard to one’s choice, and (g) informing the P1 about the location switch that occurred in her absence (all behaviors except (g) were checked in all the tasks; (g) was checked only in the two communicative tasks). If a participant displayed behavior included in the checklist while making her/his choice, the child was coded as “uncertain” for that task. The experimenters did not respond to participants’ expressions of uncertainty but just repeated the critical utterances.

In the communicative tasks, many children expressed uncertainty about their choice. In the younger group, the percentage of children who exhibited at least one form of uncertain behavior in the literal and nonliteral tasks was 38.9% and 50.0%, respectively; in the middle-aged group, the percentage of children who exhibited such behavior in the literal and nonliteral tasks was 55.0% and 35.0%, respectively; and in the older group, the number of children who exhibited such behavior in the literal and nonliteral tasks was 30.0% and 25.0%, respectively. In contrast, in the standard tasks, none of the children expressed uncertainty. As mentioned above, the children’s expressions of uncertainty indicate that they are sensitive to the self-other difference with regard to knowledge states. Thus, I categorized the children who made a correct choice or expressed uncertainty as “sensitive,” and analyzed the proportion of sensitive responses. Table 3 shows the proportions of sensitive responses of the children from each age group in the four types of false belief tasks.

Between-task comparisons revealed that the proportion of sensitive responses differed significantly in the younger and older groups (younger group: \( Q(3) = 13.76, p < .01 \); older group: \( Q(3) = 22.76, p < .001 \)). In the younger group, the proportion of sensitive responses in the nonliteral task was significantly higher than that in the two standard tasks (\( p < .05 \)). The result for the older group was the reverse: The sensitive responses in the two communicative tasks were less than those in the two standard tasks (\( p < .05 \)). No significant difference was observed in the between-task comparisons in the middle-aged group or in the between-group comparisons in the two communicative tasks (\( ps > .05 \)).

There was no significant difference in the proportion of correct and sensitive responses between the literal and nonliteral tasks across all the age groups. However, in the nonliteral task, 34.5% of the children attempted to search for the object in the empty container that had been referred to. That is, they tried to find the object in the container that was referred to, although in reality, the objects were right before their eyes, outside the containers. This fact may demonstrate that the nature of the nonliteral task was not enough to reduce the processing effort involved in inhibiting the literal meaning. It also demonstrated that children have a strong tendency not to doubt the literal meaning of adults’ utterances.

**Discussion**

The analysis of the proportion of correct responses does not support the hypothesis that communication assists children’s mind-reading. On the whole, children committed more errors in the communicative tasks than in the standard ones. On the other hand, given the high proportion of sensitive responses, younger children were more sensitive to the self-other difference with regard to knowledge states in the communicative settings than in the hypothetical situations. We cannot conclude that younger children are better at representing others’ false belief in communication because their uncertainty may merely be the result of an appreciation of the experimenter’s relative ignorance, or understanding of the “seeing is knowing” rule. That is, the children might express uncertainty about their choice not because they appreciated P1’s false belief but because they merely understood that there were differences with regard to knowledge states between them and P1, who did not see the location switch. Moreover, developmentally, the understanding of relative ignorance emerges earlier than that of false belief (Wellman & Liu, 2002). Thus, in this research,

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Sally-Ann</th>
<th>Matched standard</th>
<th>Matched literal</th>
<th>Matched nonliteral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger (n = 18)</td>
<td>22.2%</td>
<td>22.2%</td>
<td>55.6%</td>
<td>72.2%</td>
</tr>
<tr>
<td>Middle-aged (n = 20)</td>
<td>85.0%</td>
<td>45.0%</td>
<td>65.0%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Older (n = 20)</td>
<td>100%</td>
<td>80.0%</td>
<td>45.0%</td>
<td>35.0%</td>
</tr>
</tbody>
</table>

*Note.* In the Sally-Ann task and matched standard task, the proportions of sensitive responses were equal to the proportions of correct responses, because no one expressed uncertainty in these tasks.
whether communication helps young children’s false belief understanding remains unclear. However, a weaker version of the hypothesis—younger children are more eager to read the mind of the interlocutor than the hypothetical story character—is still a viable possibility. As Carpenter et al. (2002) claimed, determining adults’ communicative intentions could put children in the mindset of mind-reading.

This study also investigates the effect of the inhibition of literal meaning on performance in a communicative task. For this purpose, two types of communicative tasks—literal and nonliteral—were conducted. The number of correct and sensitive responses in the two communicative tasks did not differ significantly. However, children’s search behavior in the empty container in the nonliteral task demonstrated that when interpreting adults’ utterances, young children excessively emphasize the literal meaning of the utterances. The results showed that children sometimes believe not their own perception and rationality but what is said by adults’ utterances. This fact suggests that the inhibition of literal meaning is one of the potential causes of children experiencing difficulty in communicative false belief tasks and, possibly, in everyday verbal communication.

The performances of the middle-aged and older groups are controversial. Although a majority of these children passed the standard tasks, they failed to select the intended object in the communicative tasks. These results might indicate that there is a kind of U-shaped developmental trajectory; in other words, when interpreting the other’s utterances, the older children placed greater emphasis on the literal meaning of the utterance than younger children and adults did. However, these results might have been caused by the methodological problems in the communicative tasks. Although the settings of the communicative tasks were naturalistic until the test questions were asked, the situation changed when the children made their choice. Since the experimenter could not reveal the answers to the questions, the adults’ responses to the children’s expressions of uncertainty might have become rather unnatural. Older children might be more sensitive to this unnaturalness than younger children, and this may account for their relatively poor performance in the communicative tasks. These possibilities need to be addressed in future research.

Again, understanding others’ mental states essentially refers to our ability to communicate our intentions, desires, and beliefs to one another. Successful verbal communication is based on continuous perspective taking and shifting on the part of the participants of the conversation, that is, mind-reading ability. It is plausible that such a conversational experience plays an important role in the development of children’s mind-reading ability (e.g., de Rosnay & Hughes, 2006). If so, investigating children’s mind-reading development in communicative situations could make an important contribution to the studies of theory of mind. Recently, studies in language pragmatics have focused on the role of mind-reading in human communication (e.g., Sperber & Wilson, 1995). Thus, the investigation of the developmental change in mind-reading ability in communication can also contribute to the theoretical progress in language pragmatics, and future researches must be conducted in an interdisciplinary manner.

In conclusion, this study revealed that young children are more eager to read the mind of the interlocutor than the hypothetical story character, although this only holds true for the younger group. Do children’s expressions of uncertainty indicate their understanding of false belief? What is the exact nature of the effect of literalness on children’s communication? These questions remain for future researches. Easy or not easy—the hypothesis that young children are better at representing false beliefs in communicative settings must be tested further in more naturalistic situations, in greater detail, and under more varied conditions.

Acknowledgments

The author is grateful to Etsuko Haryu for her helpful comments during the planning of this study. This study was supported by a grant from the Japanese Society for the Promotion of Science.

References