This study examined young children’s understanding of second-order mental states, which have embedded structures such as beliefs about beliefs. Participants were 4-, 5-, and 6-year-olds. First-order and second-order false belief tasks were prepared with similar structures to minimize the difference between them. The results showed that over half of the 6-year-olds were successful on both tasks, while the 5-year-olds were successful on the first-order but not the second-order false belief task. This time lag between understanding first-order and second-order mental states is consistent with the findings of previous studies. These findings suggest that information processing of embedded mental states is a significant factor for understanding second-order mental states.

Key words: understanding of second-order mental states, second-order false belief task, young children

Humans usually understand not only mental states such as beliefs of others, but also complex mental states which have embedded structures such as beliefs about beliefs. These two types are called the understanding of first-order mental states and second-order mental states, respectively. In contrast, understanding based only on real events is called zero-order reasoning (e.g., Perner, 1988; Perner & Wimmer, 1985). Although understanding first-order mental states is important because it can be used for the simple prediction of human behavior, understanding second-order mental states is more important because it can capture more complex social interactions. Therefore, the latter’s importance has been emphasized by theorists in several fields, such as communication theory (e.g., Grice, 1975; Sperber, 1994) or game theory (e.g., Rapoport, 1967; Chwe, 2001).

Miller, Kessel, and Flavell (1970) investigated elementary school children’s ability to verbally describe embedded “thought-bubble” cartoons that corresponded to second-order mental states. They found that even 6th-graders had difficulty verbally describing these states. Subsequent studies generally confirmed the results of Miller et al. (e.g., Barenboim, 1978; Landry & Lyons-Ruth, 1980; Oppenheimer, 1986; Veith, 1980). The results were similar for studies using a reversed task, where children were given a
description and had to simply point to the corresponding thought-bubble cartoon (e.g., Eliot, Lovell, Dayton, & McGrady, 1979).

Elementary school children understand that thought-bubbles show what the character is thinking (Wellman, Hollander, & Schult, 1996), so this is not the source of the difficulty. The difficulty may be related to the complexity of the verbal task, which requires dealing with complex syntactic rules. Furthermore, the thought-bubble paradigm has the fatal flaw that correct responses do not demonstrate “understanding” of second-order mental states beyond the syntactic exercise of matching embedded sentences with embedded thought-bubbles (Perner & Wimmer, 1985).

Wimmer and Perner (1983) and Baron-Cohen, Leslie, and Frith (1985) introduced the “false belief task” to investigate the understanding of first-order mental states. Perner and Wimmer (1985) developed this task into the “second-order false belief task” to investigate the understanding of second-order mental states. The second-order false belief task has the following fundamental structure.

(a) Both actors \( A \) and \( B \) have information \( x \).
(b) While \( A \) doesn’t know, \( B \) gets the information that \( x \) changes to \( y \).
(c) While \( B \) doesn’t know, \( A \) also gets the information that \( x \) changes to \( y \).
(d) The information changes from \( x \) to \( y \).

Test question: “Which information does \( B \) think \( A \) has, \( x \) or \( y \)?”

At first, the information (zero-order), \( A \)’s first-order belief about the information, \( B \)’s first-order belief about the information, and \( B \)’s second-order belief about \( A \)’s belief about the information all have the value of \( x \). Subsequently, the first three items of information are changed from \( x \) to \( y \). Therefore, the participant must understand \( B \)’s second-order belief about \( A \)’s belief to correctly answer “\( x \)” to the test question. Any of the other strategies based on lower-order level reasoning lead to wrong answer “\( y \)” Thus, this task is better than the previous tasks using thought-bubble cartoons, because it does not require complex verbal utterances.

Based on the above structure, Perner and Wimmer (1985) produced the “Ice-Cream story” that has social interaction among four characters and showed that children understand second-order beliefs at about the age of 7 to 9. Sullivan, Zaitchik, and Tager-Flusberg (1994) reasoned that the information-processing demands of the Ice-Cream story were high for young children, and so they introduced the “Birthday Puppy story” involving fewer characters and locations and including a deception context. The performance of the 6-year-old children was better on their task.

In theory-of-mind research, there have been a lot of studies and fruitful findings about the understanding of first-order mental states. These findings generated meta-analysis studies for the first-order false belief tasks (e.g., Goushiki, 2000; Wellman, Cross Watson, 2001). Children usually understand the first-order false beliefs by about the age of 4 or 5 (e.g., Wellman et al., 2001). In contrast, there have not been many empirical studies about children’s understanding of second-order mental states. Therefore, even the age when children understand second-order mental states has not been completely
established. But, it may be so far concluded from previous findings that it is difficult to understand second-order beliefs until at least about the age of 6, and that there is at least a 1 to 2-year lag between the understanding of first-order and second-order beliefs (e.g., Hasselhorn, Mähler, & Grube, 2005; Sullivan et al., 1994).

The amount of information processing needed to understand second-order beliefs should be more than for first-order beliefs, because the former structure includes the latter structure recursively. However, the previous second-order false belief tasks (the Ice-Cream story and the Birthday Puppy story) were too different from the first-order false belief tasks in story and format to allow for direct comparisons minimizing the confounding factor of different information processing demands. Therefore, these factors of unnecessary complexity could have influenced the performance in these second-order false belief tasks, so that it is difficult to be confident about the previous findings of the age when children understand second-order mental states and the length of the lag between the understanding of first-order and second-order beliefs.

A better way to solve these issues is to match the first-order and second-order false belief tasks as closely as possible and to compare them directly using the same participants. In fact, the simplified second-order false belief format, transformed from the standard first-order false belief task, has recently been used (e.g., Astington, Pelletier, & Homer, 2002; Hayashi, 2002; Sodian, & Hülsken, 2005). However, in some studies such as Astington et al. (2002) the tasks included extra elements, for instance that there were four characters in the story despite that two characters were enough. Hayashi (2002) adapted the simplified second-order false belief format to the standard first-order false belief task so that the only difference between them is the difference between first-order and second-order mental states. He compared three types of second-order false belief tasks (the Ice-cream story, the Birthday Puppy story, and the simplified task) and the standard first-order false belief task within the same children, about 7 through 12 years old. The results showed that the Ice-Cream story and the Birthday-Puppy story had unnecessary complexity, whereas the simplified task did not have this problem. However, it was not possible to confirm the age when children understand second-order mental states and to what extent there is a time lag between the understanding of first-order and second-order mental states, because most 7-year-olds gave correct answers for the simplified task as well as the standard first-order false belief task.

Therefore, it is necessary to compare the simplified task and the standard first-order false belief task directly using the same younger children. The first purpose of the present study is to confirm the age when children understand second-order mental states by using the simplified task. The second purpose is to investigate to what extent there is a time lag between understanding first-order and second-order mental states. Using this simplified task is important, because it allows us to judge whether the specific information processing required for handling embedded mental states recursively is relatively important for understanding of second-order beliefs.

A time lag could be caused by the specific information processing required for handling embedded mental states and other information processing required for dealing with unnecessary complexity. The former is common for any type of second-order false
tasks, whereas the latter would be reduced by using the simplified task because this task is set up to be the same as the standard false belief task except for handling the order of mental states. Therefore, if the time lag decreases greatly, then this would mean that the previous findings about the time-lag were due to unnecessary task story complexity, and that information processing related to embedded mental states does not have a significant influence on understanding of second-order mental states. In contrast, if the time lag remains as great as in the previous findings, then this would mean that information processing of embedded mental states has a significant influence.

**METHOD**

*Participants*
Fifty-three Japanese children participated in the experiment. Data from two children were excluded because of their failure to pay attention. The sample consisted of 15 4-year-olds (10 boys and 5 girls; mean age 4:7 years; range 4:4–4:11), 23 5-year-olds (9 boys and 14 girls; mean age 5:6 years; range 5:0–5:11), and 13 6-year-olds (5 boys and 8 girls; mean age 6:3 years; range 6:0–6:7). All attended kindergarten in the Osaka area.

*Materials*
The first-order and the second-order false belief tasks involved four cartoon-style pictures. Two stories were prepared: the Chocolate story and the Doll story. One was assigned for the first-order false task and the other was assigned for the second-order false belief task. This assignment was counterbalanced between participants. Table 1 shows two types of tasks using the Chocolate story1. The only difference was whether it was described that “the girl looks at the boy’s action from outside the window, but the boy is not aware of this” in both the cartoon and the sentences in Episode 3.

*Design*
A within-subjects design was used for the first-order false belief and second-order false belief conditions. Half completed the first-order false belief task before the second-order false belief task, and the other half had the reverse order.

*Procedure*
Each participant was tested individually. The experimenter followed a routine script as shown in Table 1. Each participant was first given a name labeling pre-test. The experimenter showed two cartoons, one depicting a refrigerator, a basket, and a window in a room, and the other only depicting the chocolate. After the experimenter named all of them, the experimenter pointed each of them and asked the participant, “What is this?” This was the qualifying condition for acceptance in the experiment. All participants passed this pre-test.

Following Leekam and Perner (1991), the tasks all ended with a series of questions. First a reality question, then a memory question, and finally a test question as shown in Table 1. Only in the second-order false belief task did the experimenter asked for a justification of the answer to the test question. If participants gave wrong answers to the reality question or the memory question, they were told the story again. If they still failed to respond correctly, the answer was given to them.

---

1 In the Doll story, x, y, and a were a box, a bag, and a doll, respectively. In order to prevent the children from confusing two similar tasks, the moving direction of the doll and the role of a boy and a girl were replaced.
Preliminary analysis

For the first-order false belief task, one participant was told the story again because he gave the wrong answer to the memory question, but did not need to be given the answer. For the second-order false belief task, 14 other participants were told the story again, and six of them needed to be given the answer. One other participant was told the story again in both tasks, but did not need to be given the answer. Preliminary analysis indicated that there was no significant difference between the two stories, between the presentation order of the two types of tasks, or between sexes in the performance of each task, so data for these factors were combined.

Test question

The numbers of children giving correct answers to the test question are shown in Table 2. For the first-order false belief task, the percentages of correct answers among the 4-, 5- and 6-year-old children were 40%, 78%, and 85%, respectively. The 5- and 6-year-old children were responding at above chance levels, while 4-year-olds were not responding at above chance levels. A chi-square test revealed an overall effect of age ($\chi^2(2) = 8.24, p < .05$). This occurred because significantly fewer of the 4-year-olds passed than did the 5- and 6-year-olds ($\chi^2(1) = 5.71, p < .05$; $\chi^2(1) = 5.81, p < .05$). For
the second-order false belief task, the percentages of correct answers of 4-, 5-, and 6-year-olds were 20%, 35%, and 69%, respectively. The 5- and 6-year-olds were not responding at above chance levels, while the 4-year-olds were responding at under chance levels. A chi-square test revealed an overall effect of age ($\chi^2(2) = 7.43, p < .05$). This occurred because significantly more of the 6-year-olds passed than did the 4- and 5-year-olds ($\chi^2(1) = 6.89, p < .01; \chi^2(1) = 3.95, p < .05)^2$.

As shown in Table 2, 16 children gave a correct answer to the first-order false belief task but an incorrect answer to the second-order false belief task, while only one child showed the opposite response pattern (McNemar’s test: $\chi^2(1) = 13.2, p < .001)^2$.

### Justifications

The children’s justifications on the second-order false belief task were coded into the following categories, based on Perner and Wimmer (1985) and Sullivan et al. (1994).

**Explicit second-order reasoning:** One character’s mental state embedded in the other character’s mental state was mentioned, e.g., “Because the boy doesn’t know that the girl knows where the chocolate really is.”

**Implicit second-order reasoning:** Relevant information embedded in one character’s mental state was mentioned, e.g., “Because the boy doesn’t know that the girl looked at his action.”

**Initial location:** The original location of the chocolate was mentioned, e.g., “Because the girl put the chocolate into the refrigerator in the beginning.”

**First-order reasoning:** One character’s mental state was mentioned, e.g., “Because the boy knows that the chocolate is in the basket.”

**Zero-order reasoning:** No character’s mental state was mentioned, but the object’s current location was mentioned, e.g., “Because the chocolate is in the basket.”

**Story facts:** No character’s mental state was mentioned, but facts were mentioned,

---

2 Even if six children needed to be given the answer for the reality question or the memory question were removed, these results remained almost the same.
e.g., “Because the girl wants to eat the chocolate.”

*Nonsense*: Nonsensical information was mentioned.

*No response*: The participants said that they didn’t know or said nothing.

Table 3 shows the frequency of justifications of correct and wrong answers to the test question. There was no explicit second-order reasoning and first-order reasoning in the present study sample. Following Perner and Wimmer (1985) and Sullivan et al. (1994), implicit second-order reasoning, and initial location are appropriate justifications, and the rest of the categories are inappropriate justifications. As shown in Table 3, 70% of participants who answered the test question correctly gave appropriate justifications. In contrast, all of participants who answered the test question incorrectly gave inappropriate justifications².

### DISCUSSION

The first purpose of the present study was to confirm the age when children understand second-order mental states by using the simplified second-order false belief task. Although over half of the children passed the simplified task at about the age of 6, their performance was not above chance level. Hayashi (2002) used almost the same tasks as the present study with elementary school children, and showed that the percentages of correct answers were 90% to 100%. Therefore, based upon the results of the present study and Hayashi (2002), it is concluded that Japanese children come to understand second-order mental states at around the age of 6 to 7. This result generally corresponds to the other findings using a simplified task format such as Astington et al. (2002), which showed that 6- to 7-year olds passed a similar simplified task, but the younger children did
not pass. Furthermore, this result generally corresponds to findings for the Birthday Puppy story, such as Sullivan et al. (1994).

The second purpose of the present study was to investigate to what extent there is a time lag between understanding first-order and second-order mental states. Most children explicitly passed the first-order false belief task at about the age of 5. This result corresponds to the previous findings for Japanese children (Wellman et al., 2001). As described above, the children passed the simplified second-order false belief task at around the age of 6 to 7. Despite that the only difference between the two types of tasks was the difference between first-order and second-order mental states, these results show that there was a time lag between understanding of first-order and second-order mental states comparable to the previous findings. These findings mean that information processing of embedded mental states has a significant influence for understanding of second-order beliefs, and that it takes 1 or 2 more years after understanding first-order beliefs to consistently perform this specific information processing.

Overall, the children’s performance did not improve dramatically, such as showing a large decrease in the time lag, and was almost the same as in previous studies. Rather, no child was able to produce explicit second-order reasoning for justification, whereas there have been a few cases in the literature that even four and five year-old children have produced this type of reasoning (Sullivan et al., 1994). In addition, the proportion of implicit second-order reasoning for appropriate justifications was less than in previous studies.

The children’s performance in the present study may be somewhat lower than in other studies due to the cultural difference. Wellman et al. (2001) demonstrated in their meta-analysis that the country of origin influenced performance. They showed that, although the development follows a similar age trajectory across many cultures, children in Japan performed somewhat worse on the first-order false belief tasks than children in the United States and the United Kingdom. This may also apply to second-order mental states, namely, Japanese children may perform somewhat worse on the second-order false belief tasks than Western children, especially for justifications.

Further studies are needed to follow up on the findings of the present study that the specific information processing of embedded mental states has a significant influence.

The first future direction is to investigate whether children become able to perform the specific information processing of embedded mental states as a function of a conceptual reorganization or simply as a function of increasing general information processing abilities. The former hypothesis is proposed by Perner (1988) that “second-order attributions require understanding of the recursive nature of mental states” involving a conceptual reorganization. The latter hypothesis is proposed by Sullivan and her colleagues that “once children understand the representational nature of mental states ..., no further conceptual development is needed to recursively embed mental states” (Sullivan et al., 1994; Tager-Flusberg & Sullivan, 1994). Recent studies have investigated the relationship between working memory ability or inhibition control and the understanding of not only first-order mental states (e.g., Carlson, Moses, & Breton, 2002) but also second-order mental states (e.g., Hasselhorn et al., 2005; Perner, Kain, &
Barchfeld, 2002; Sodian & Hülsken, 2005). For example, Hasselhorn et al. (2005) showed that the understanding of second-order beliefs was significantly related to phonological working memory and language abilities. They provided a hypothetical model of the developmental dependencies between them, which posits that the phonological working memory capacity constrains the development of verbal abilities, and these abilities develop the understanding of first-order and second-order mental states. However, little is known about whether specific information processing of embedded mental states occurs by means of a kind of conceptual reorganization.

The second future direction is to investigate the contents of the specific information processing of embedded mental states in detail. The question here is whether the difficulty in understanding second-order beliefs is due to the difficulty in understanding recursive states in general or recursive representational states in particular. If the difficulty is in understanding recursive states in general, we have to investigate the ability of children to deal with recursive structures in different areas, such as language. Generally, language has a recursive structure, for example, a sentence can include a relative clause in the middle of the same kind of relative clause (Pinker, 1994). In fact, the empirical study of dealing with recursive structures has recently been started (e.g., Tokimoto & Okanoya, 2004). On the other hand, if the difficulty is particular to recursive representational states, this leads to a new question: Is the difficulty understanding second-order beliefs due to a difficulty understanding recursive representational mental states in particular or not?

Zaitchik (1990) used a first-order belief task and a first-order photo task that demonstrated a difference of understanding between mental representations and non-mental representations. If we apply this framework to second-order states, we can investigate whether there is a difference in difficulty between the understanding of “A person thinks that the other person thinks ...” and the understanding of “A photo shows that the other photo shows ...” Here, children need to understand the recursive representational mental states in the former and only the recursive representational states in the latter. If the difficulty in understanding second-order beliefs is due to a difficulty with recursive representational mental states in particular, then the former should be more difficult than the latter.

Finally, in order to examine the significance of understanding of second-order mental states in communication, it is also important to investigate not only understanding of second-order mental states but also the relationship between this understanding and other social cognitive abilities. Previous studies showed that understanding second-order mental states was correlated with responsibility attribution (Yuill & Perner, 1987), distinguishing a lie from a joke (e.g., Hayashi, 2002; Leekam, 1991; Sullivan, Winner, & Hopfield, 1995; Winner, Brownell, Happé, Blum, & Pincus, 1998) or irony (Winner & Leekam, 1991), predicting self-presentational facial displays (e.g., Banerjee & Yuill, 1999), and understanding white lies (Broomfield, Robinson, & Robinson, 2002). Theoretical viewpoints such as Relevance theory (Sperber & Wilson, 1995) would be also important. This theory explicates the role of comprehension of intentions in human communication, and suggests that there are two types of intentions: informative intention
and communicative intention. Thinking that someone has an informative, or a communicative intention, requires understanding second- and higher-order mental states (Sperber, 1994). It is recommended that further studies should be undertaken to investigate this theory.

REFERENCES


(Manuscript received March 4, 2006; Revision accepted April 28, 2006)