

# How Do Young Children Process Beliefs About Beliefs?: Evidence from Response Latency

HARUO KIKUNO, PETER MITCHELL AND FENJA ZIEGLER

---

**Abstract:** Are incorrect judgments on false belief tasks better explained within the framework of a conceptual change theory or a bias theory? Conceptual change theory posits a change in the form of reasoning from 3 to 4 years old while bias theory posits that *processing* factors are responsible for errors among younger children. The results from three experiments showed that children who failed a test of false belief took as long to respond as those who passed, and both groups of children took longer to respond to belief questions than to questions about prior states of reality. These results seem to support the bias theory.

## 1. Introduction—From Failing to Passing Tests of False Belief

Wimmer and Perner (1983) devised the seminal method for testing understanding of false belief by presenting children with an unexpected transfer task. The following is an abridged example, which is normally illustrated with pictures or handpuppets: Maxi eats some chocolate and puts the rest in the green cupboard. He goes out to play. In the meantime his mother uses some of the chocolate to make a cake. She puts the remaining chocolate into the blue cupboard. Children are then asked where Maxi will look for his chocolate. In order to establish that children do not just chance upon the correct response, they also answer two control questions. The memory control asks where Maxi put the chocolate and the reality control asks about the current state of reality, that is, where the chocolate is at the present time.

Many studies report that children younger than four years fail to acknowledge false belief (Wellman, Cross and Watson, 2001) even if they answer the control questions correctly. Some explain failures by suggesting that young children lack a *concept* of belief (e.g. Gopnik, 1993; Perner, 1991) or have an incomplete concept (e.g. Wellman, 1990), while others propose that *processing* factors lead to errors early in development (e.g. Fodor, 1992; Frye, Zelazo and Palfai, 1995; Gordon and Olson, 1998; Leslie and Thaiss, 1992; Mitchell, 1996; Riggs, Peterson, Robinson and Mitchell, 1998; Roth and Leslie, 1998; Zelazo, 2000). The purpose of this study is to investigate how long it takes children to respond to questions about beliefs and factual information as a way of testing processing accounts against accounts that propose conceptual change. Below, we consider what rival theories say about the basis of children's errors in a test of false belief.

---

**Address for correspondence:** School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD

**Email:** peter.mitchell@nottingham.ac.uk

*Mind & Language*, Vol. 22 No. 3 June 2007, pp. 297–316.

© 2007 The Authors

Journal compilation © 2007 Blackwell Publishing Ltd

### 1.1 Conceptual Change Theory

Following Dennett's (1978) commentary on Premack and Woodruff's (1978) investigation into a chimpanzee's sensitivity to a human actor's desire, Wimmer and Perner (1983) specified the minimal conditions under which we should credit children with a theory of mind: Being able to acknowledge false belief. After conducting a meta-analysis, Wellman *et al.* (2001) reported that children aged 4 years stand a better chance of successfully acknowledging false belief than children younger than 4 years. They, along with others (e.g. Gopnik, 1993), concluded that around the time of their fourth birthday, children undergo a radical shift that culminates in the acquisition of a concept of belief. In drawing this conclusion, researchers (e.g. Gopnik and Astington, 1988; Wellman *et al.*, 2001) stressed that there is a sharp change in performance at the age of 4 years.

In a sense, this is an 'obvious' explanation for the developmental trend: When working out where Maxi will look, you either take his false belief into account or you do not. If you take it into account, then you deserve to be credited with a theory of mind—at least according to Dennett's (1978) definition; if you do not take Maxi's false belief into account, then perhaps it is a sign that you lack a concept of belief. An assumption being made here is that you would use a concept if you possessed it.

The trouble with the conceptual change account is manifold. In principle, it does not necessarily follow that you would use a concept (or use it correctly) just because you possess it: There is a difference between competence and performance. On empirical grounds, the developmental trend is conspicuously not sharp. Contrary to the text in Wellman *et al.*'s (2001) and Gopnik and Astington's (1988) articles, it is clear from the plots of performance against age shown in those same articles that there is a gradual improvement. This is true not just in cross-sectional comparisons but also in longitudinal data (Flynn, O'Malley and Wood, 2004).

Besides, any account of errors in a test of false belief needs to be able to explain why there is a systematic tendency to report reality among young children. If children lacked a concept of belief, then by default presumably they would answer in a random and unsystematic way when asked what another person is thinking—in the above example children would predict roughly half the time that Maxi will search in the green cupboard and half the time they would predict his search in the blue cupboard. In fact, though, children aged around 3 years tend to give systematically incorrect judgments by reporting that Maxi will search in the place where the chocolate is currently located. Wimmer and Hartl (1991) offered a neat explanation for systematic errors from a conceptual change perspective. They proposed that if a child lacked a concept of belief, they would be constrained to process the test question as if it did not contain explicit or implicit reference to belief. So children would be conceptually constrained to interpret, 'Where will Maxi look for the chocolate?' as, 'Where is the chocolate?' and hence systematically report the chocolate's current location.

## 1.2 Processing Theories

Bias theory is a processing account, which says that children make errors in a test of false belief not because they lack the relevant concept but because they are prone to systematic error when calculating belief. This comes into focus when considering that even adults systematically confuse their own belief with other people's in special circumstances (Keysar, Lin and Barr, 2003; Mitchell, Robinson, Isaacs and Nye, 1996). Presumably, this does not warrant the conclusion that adults lack a concept of belief; rather it seems that sometimes they are prone to bias when calculating belief. By the same token it is questionable whether errors on a test of false belief among children aged 3 years signals that they lack a concept of belief (*pace* e.g. Gopnik, 1993; Perner, 1991; Wimmer and Perner, 1983).

Several theories postulate that children's errors in a test of false belief stem from a problem in processing information (e.g. Fodor, 1992; Frye *et al.*, 1995; Leslie and Thaiss, 1992). We shall focus on a processing account called the reality bias theory (Mitchell, 1996), which proposes that although children have a concept of belief, they make a systematic error in calculation that leads them to report the current state of reality as the content of the belief. In other words, it is proposed that they do the right kind of processing for making a judgment of belief, but get the content wrong (see also Leslie and Thaiss, 1992 for a similar account). This view is enlightened by discoveries and theories of the way children reason about beliefs. Riggs *et al.* (1998) found a very high correlation between children's success in counterfactual reasoning and their success in acknowledging false belief. In the counterfactual task, children were presented with the story about Maxi and his chocolate, based on Wimmer and Perner (1983), and then were asked the question, 'If Mum had not baked the cake, where would the chocolate be now?' In order to tackle this question we first need to imagine that Mum did not bake a cake and then consider the chocolate's location in that imaginary world. According to Peterson and Riggs (1999), we need to follow the same procedure when answering a question about belief, except that we take an additional step of using the output of our counterfactual reasoning for attributing a belief. For example, in the mental world that we imagine Maxi inhabits, Mum did not bake a cake and therefore she did not move the chocolate.

Two related implications arising from the counterfactual reasoning account are especially noteworthy. First, individuals are prompted to carry out the requisite counterfactual reasoning by the test question enquiring about Maxi's belief. If the test question (or some other relevant prompt) had not been presented, then individuals would not necessarily calculate Maxi's belief spontaneously (Peterson and Riggs, 1999). That is, individuals are not simply able to *retrieve* the answer from their knowledge base, at least not one that relates to belief; rather they *calculate* a response on demand. This distinctive feature of processing is not considered by other bias theories, such as Alan Leslie's. Second, and in consequence, the time an individual takes to respond to the belief question is bound to reflect the time it took them to carry out the necessary reasoning. In contrast, answering a question about the prior state (where the chocolate was initially) or the current state (where the chocolate is now) merely depends on retrieval and does not entail any

calculation or formulation. Hence, the particular bias theory we investigate in this article makes unique predictions.

One aspect of calculating a belief deserves special mention. Perhaps the principal challenge faced by a child (Mitchell and Lacohee, 1991) or adult (Mitchell *et al.*, 1996) surrounds the conflict that needs to be resolved when making a judgment of *false* belief. In this case, the participant has to set aside conflicting knowledge they possess about the true state of reality (Harris, 1991) as in, 'Maxi only saw some of the things that I saw, so from Maxi's perspective Mum had not baked a cake. In that case, will he think the chocolate is in the green cupboard or the blue cupboard? The chocolate is in the blue cupboard but wait, Maxi will think (falsely) that it is in the *green* cupboard.' If a participant did not possess a concept of belief, and failing a test of false belief was a sign to that effect (as proposed by conceptual change theorists), then presumably participants would not experience the kind of representational conflict suggested above and neither would they engage in counterfactual reasoning as in, 'The chocolate is in the blue cupboard so the answer is 'blue cupboard.' If this is correct, then all things equal, false belief passers would take longer to respond than false belief failers, because only the passers experienced the time-consuming counterfactual reasoning and representational conflict.

But suppose that failing a test of false belief is not a sign of lacking the concept of belief, contrary to conceptual change theorists, and moreover that false belief failers do engage in counterfactual reasoning and do experience time-consuming representational conflict as in, 'Maxi only saw some of the things that I saw, so from Maxi's perspective Mum had not baked a cake. In that case, will he think the chocolate is in the green cupboard or the blue cupboard? The chocolate is in the blue cupboard and so Maxi will think that it is in the *blue* cupboard.' According to this suggestion, the passer and the failer execute the same kind of reasoning and differ only in the content of their output. Perhaps the main difference between the passer and the failer is that the bias towards current reality is stronger in the failer, leading to a higher likelihood of error at the stage of output selection (c.f. Leslie and Thaiss, 1992). Note that much of human reasoning is subject to bias (e.g. Nisbett, 2003), but this does not necessarily signify the lack of a concept.

In short, we assume that young children engage the appropriate form of reasoning but make errors of content (cf. Roth and Leslie, 1998). It follows, therefore, that offering support to the correct content should lead to improved performance (cf. Mitchell and Lacohee, 1991; Saltmarsh and Mitchell, 1998; Saltmarsh, Mitchell and Robinson, 1995). In view of that possibility, a subsidiary component of the studies involved pitting a condition in which children had support in identifying the correct belief content ('Aspect' condition) against conditions where support was not so strong ('Discriminative' and 'Standard' conditions).

### 1.3 Rationale and Hypotheses

The standard condition is based on the unexpected transfer task, and describes a boy, Taro, playing with a Plasticine hat that he places in Location 1. On leaving

the scene, his sister, Hanako, appears and moves the Plasticine hat to Location 2. Children are then asked where Taro will look for the Plasticine. In the Aspect and Discriminative conditions the story is slightly changed, so that the object that is played with changes before it is transferred to the new location; the conditions differ in the questions children are asked. In both, Taro put a Plasticine hat into Location 1. Later, Hanako transformed it into a Plasticine apple and then put it into Location 2. In the Aspect condition, children were asked about the aspect of the object in its initial form ('Where will Taro look for his hat?'), whereas in the Discriminative condition, they were asked about the more general class of the object ('Where will Taro look for his Plasticine'). We expected that the Aspect condition would help children to calculate Taro's belief based on the aspect of the object in the form of a hat, which was associated with Location 1. By referring to the object as the 'hat' rather than the more generic 'Plasticine' children should be helped in elaborating on the different states of the object and the different locations that the object is associated with. The Plasticine as a 'hat' is only associated with Location 1 (the false belief location), but not with Location 2 (the actual location of the Plasticine). The general class of the object is associated with Locations 1 and 2, in the sense that the Plasticine (as a hat or an apple) resided in both places at some point in the procedure. The more specific reference of the Aspect test question (hat) should assist children in identifying Location 1 as the content of Taro's belief, given that his belief concerns the Plasticine as a hat, which was uniquely identified with Location 1 in this procedure (see also Mitchell and Kikuno, 2000).

Different predictions about performance in these conditions can be derived from the bias theory and the conceptual change theory. The bias theory predicts that the Aspect condition will help children to acknowledge false belief, by reducing the level of representational conflict, and this might be measurable as a shortened response time as well as an increase in the incidence of correct judgments. The facilitation in correct responding is something that might be predicted by other theories as well (e.g. Leslie's and Fodor's) but the predictions about response time is beyond the scope of those theories. Performance in the other conditions would be inferior if children's limited processing led to their incorrect selection from previous and current representations as a basis for calculating Taro's belief. In contrast, the conceptual change theory might have no grounds for predicting improved performance in the Aspect condition, given that young children supposedly lack a concept of belief and therefore perhaps would not benefit from support in selecting the appropriate information in formulating a judgment of belief.

In summary, the main purpose of the studies was to measure response time to questions in a test of false belief. Bias theory makes specific predictions that are not made by certain versions of conceptual change theory. Experiment 1 offered a basic test of these predictions, while Experiment 2 offered a replication with a different procedure. Experiment 3 included a test of true belief to examine the prediction that response time was faster on this than on a test of false belief.

If response time was not faster in a test of true belief, then it would be impossible to rule out difficulties with question comprehension, as a rival explanation for long response time. A more detailed rationale is provided in the introduction to Experiment 3.

## 2. Experiment 1

### 2.1 Method

**2.1.1 Participants.** One hundred and fourteen children (59 boys and 55 girls) participated in this study. They were divided randomly into three groups, each consisting of 38 children. They ranged from 3:05 to 4:05 (mean age = 3:11). The children attended three nursery schools of Osaka and Nara prefecture in Japan situated in middle class areas.

**2.1.2 Materials.** A story about a boy named Taro was constructed, based on Wimmer and Perner's (1983) unexpected transfer task. The materials consisted of a boy doll (Taro), a girl doll (Hanako), a miniature green drawer, a miniature yellow cupboard, and a miniature wall with windows.

**2.1.3 Procedure.** This experiment consisted of three conditions that were the Standard, the Discriminative, and the Aspect conditions. Each child participated individually in one of these in a between groups design. They received the story about Taro and then were asked three questions: belief, reality, and memory. The story and the questions were different among conditions (see below).

Children in the Standard condition were presented with the following story. Taro made his Plasticine in the shape of a hat. He decided to play with his Plasticine later, so he deposited it in the green drawer. He left the room to play outside. In his absence, his sister who is called Hanako appeared. She found Taro's Plasticine in the green drawer, so she got it out of the green drawer and she played with his Plasticine. Then she put the Plasticine (still in the shape of a hat) in a yellow cupboard. Subsequently, Taro was returning.

Participants were then asked 3 questions in random order (determined by a random number generator on a calculator), the critical belief question and two control questions:

Belief question:	Where will Taro look for his Plasticine?
Memory question:	Where is his Plasticine really?
Reality question:	Do you remember where Taro put the Plasticine in the beginning?

Children in the Discriminative condition were presented with the following Discriminative story. Taro made his Plasticine in the shape of a hat. He decided to play with his hat later, so he deposited it in the green drawer. He left the room to play outside. In his absence, his sister who is called Hanako appeared. She found Taro's Plasticine in the green drawer, so she got it out of the green drawer and she

played with his Plasticine to make an apple. Then she put the Plasticine in the shape of an apple in a yellow cupboard.

Children in the Discriminative condition received the same questions as those in the Standard condition. Those in the Aspect condition were asked different belief and memory questions, where 'Plasticine' was replaced by 'hat'.

Response times were recorded by a stopwatch, which was operated discretely under the table out of children's sight. The experimenter told children that they should respond correctly rather than quickly. We defined the response time as the interval between the experimenter finishing the question and the child starting to answer. The experimenter remained expressionless during questioning with the aim of minimizing cues that might lead children to respond prematurely. He looked children in the face on asking questions, but then looked down to remove any further possibility of cueing. The experimenter's recording of response time could not have been influenced by his knowledge of the correctness of response, given that the recording was completed before it was apparent whether the child had responded correctly or incorrectly.

It would have been ideal to record response times automatically or from tape recordings. However, there were two major obstacles. First, it would have been impractical to present the false belief stories on a computer capable of recording response times. Young children need to be engaged by a human experimenter in order to secure their attention on the task. Second, schools would not grant permission to tape-record the children on ethical grounds. Hence, we had no opportunity to score response times post hoc. However, this research is not unique in testing theories based on response times recorded contemporaneously by a stopwatch. For example, performance on embedded figures and block design tests are recorded in this way and the data have been used to test prominent theories relating to heightened visuo-spatial abilities in autism (Shah and Frith, 1983, 1993).

## 2.2 Results

**2.2.1 Children's Judgements.** Table 1 shows the percentage of children who answered correctly on each question. Chi-square analysis indicates that the effect of condition was significant on the belief question ( $\chi^2 = 7.12$ ,  $df = 2$ ,  $n = 114$ ,  $p < .05$ ). Paired comparisons indicate significant differences between the Standard and the Aspect condition ( $\chi^2 = 4.46$ ,  $df = 1$ ,  $n = 76$ ,  $p < .05$ ) and between the Discriminative and the Aspect condition ( $\chi^2 = 4.46$ ,  $df = 1$ ,  $n = 76$ ,  $p < .05$ ). However, the differences between conditions for the reality ( $\chi^2 = 1.88$ ,  $df = 2$ ,  $n = 76$ ) and the memory ( $\chi^2 = 0.85$ ,  $df = 2$ ,  $n = 76$ ) questions were not significant.

**2.2.2 Response time.** Because our hypotheses about response times include children who give correct and incorrect judgments, all participants were included in the reported analyses, except for those with inordinately long response times. Response times over 20 seconds were three SDs from the mean on the belief

**Table 1.** *Percentage of children who answered correctly on each question in Experiment 1.*

Condition	N	Belief	Reality	Memory
Standard	38	47.37 (18)	92.10 (35)	81.58 (31)
Discriminative	38	47.37 (18)	97.37 (37)	81.58 (31)
Aspect	38	73.68 (28)	92.10 (35)	86.84 (33)

Note. *The values in parentheses are the number of children who answered correctly.*

question and were excluded. Table 2 shows the mean response time of the children who responded correctly and incorrectly on each question. We computed a multivariate analysis of variance (MANOVA) of the design  $3 \times 2 \times 3$ : The first factor is between-subjects, which classifies the condition including the Standard, the Discriminative and the Aspect condition. The second factor is within-subjects and is the response type including correct and incorrect responses on each question. The third factor is also within-subjects and includes the belief, the memory and the reality question. The main effect of the question was significant [ $F(2, 198) = 14.91, p < .001$ ] but the other main effects and interactions did not approach significance. Paired comparisons using *t*-tests indicated that the response time on the belief question (3.91) was longer than that on the reality (1.99,  $p < .05$ ) and the memory question (2.79,  $p < .05$ ), but the difference in response time between the reality and the memory question was not significant. We repeated these analyses after excluding children who failed one or both control questions, and found no changes to any of the significant effects. The same was true in Experiments 2 and 3 (see below).

### 2.3 Discussion

More children answered correctly on the belief question in the Aspect condition than in the Standard and the Discriminative conditions. This suggests that perhaps the Aspect condition helped children to calculate their judgment of belief correctly. If so, however, this was not apparent in terms of faster response time to the belief question in the Aspect condition than in the other two conditions. Hence, the effect of the Aspect condition is either weak, limited in scope or both, suggesting perhaps that young children's bias to report current reality is rather robust.

Moving on to the discovery of focal interest, the response time on the belief question (across the three conditions) was longer than that on the memory and reality questions. The pattern of response time was not affected by whether children made correct or incorrect judgements of false belief. These results are largely consistent with the bias theory and suggest that processing on the belief question may be different from that on the memory question, even in children who give an incorrect judgment on the belief question. Importantly, children who gave an incorrect judgement on the belief question were not behaving as if they had merely been asked about the current state of reality.



**Table 2.** Mean response time on each question in Experiment 1.

Condition	N	Belief	Reality	Memory
Standard				
Correct	20	3.95 (3.77)	1.56 (2.00)	2.92 (2.42)
Incorrect	18	4.35 (4.12)	1.58 (1.40)	3.86 (3.27)
Discriminative				
Correct	18	3.92 (2.89)	2.32 (2.11)	2.82 (4.17)
Incorrect	15	4.03 (4.48)	1.45 (1.28)	2.92 (2.37)
Aspect				
Correct	25	3.29 (2.96)	2.76 (2.71)	2.06 (3.00)
Incorrect	9	4.44 (3.05)	1.86 (2.21)	2.06 (1.81)

Note. The values in parentheses are the SD.

### 3. Experiment 2

Experiment 2 employed a slightly different procedure with the principal aim of demonstrating the robustness of the findings relating to response time. In the previous experiment, a correct answer to the belief question was the same as the correct answer to the memory question. Despite that, children seemingly arrived at this correct answer via different routes, given the difference in response times. Nevertheless, in Experiment 2, the procedure was modified such that children who performed correctly gave different answers to these two questions. If children continued to take longer to answer belief than memory questions in this slightly more complicated task, it would testify to the broad scope of the effect.

#### 3.1 Method

**3.1.1 Participants.** Sixty children participated in this experiment (30 boys and 30 girls). Their age ranged from 3:07 to 4:08 (mean age of 4: 02). All attended two kindergartens of Osaka prefecture in Japan situated in a middle class area.

**3.1.2 Materials and Procedure.** Children were tested individually and each participated in the Aspect and the Discriminative condition. Children heard one story about Taro and another about Rie. The stories were the same for all children, but the belief question was different according to the condition while the reality and the memory question were constant for all children. The belief question in the Aspect condition was, 'Where will Taro look for his hat?' and in the Discriminative condition it was, 'Where will Taro look for his Plasticine?'; the correct answers to both is 'yellow cupboard. The correct answer to the memory question ('Where is the Plasticine right now?') and to the reality question ('Where was Taro's hat at the very beginning?') is 'green cupboard. Therefore, the correct answer to the belief question was different from the correct answer to memory question in the two stories (see Appendix 1 for complete stories). The stories were

presented in a fixed order, but condition (Aspect, Discriminative) varied between children in counterbalanced order.

### 3.2 Results

**3.2.1 Children's Judgements.** Sixty-eight percent ( $n=41$ ) of children responded correctly in the Aspect and 70 percent ( $n=42$ ) responded correctly in the Discriminative condition to the belief question. The percentages of children who responded correctly to the reality and the memory question were 95.00 ( $n=57$ ) and 98.33 ( $n=59$ ) in the Aspect condition and 98.33 ( $n=59$ ) and 96.67 ( $n=58$ ) in the Discriminative condition. The difference between conditions was not significant on the belief question ( $\chi^2=0.16$ ,  $df=1$ ), on the reality question ( $\chi^2=2.33$ ,  $df=1$ ), and on the memory question ( $\chi^2=0.00$ ,  $df=1$ ).

**3.2.2 Response Time.** Table 3 shows the mean response time for each condition and each question. The response time was analyzed by MANOVA of the design 2(response type: correct or incorrect)  $\times$  3 (question: belief, reality or memory) on the Aspect condition and the Discriminative condition separately, given that response type (correct or incorrect) varied between the two conditions due to the repeated measures quality of the design. The MANOVA based on the Aspect condition indicates that the main effect of the question was significant [ $F(2, 110) = 8.34$ ,  $p < .01$ ]. All other effects were nonsignificant, though the main effect of response type approached significance:  $F(1,55) = 3.91$ ,  $p < .1$  but  $p > .05$ . The MANOVA based on the Discriminative condition gave rise to a similar pattern, with a main effect associated with the question [ $F(2, 110) = 9.26$ ,  $p < .01$ ]. None of the other effects approached significance.

Because there was no effect associated with response type, we combined data from the two conditions without classifying according to response type. We computed a MANOVA with condition and question in a repeated measures design, which identified a main effect associated with the question:  $F(2, 112) = 16.50$ ,  $p < .001$ . No other effects approached significance. Paired comparisons using  $t$ -tests indicated that the response time on the belief question was significantly longer than on the reality and the memory question ( $ps < .01$ ) and that the response time on the memory question was significantly longer than on the reality question ( $p < .01$ ).

**Table 3.** Mean response time on each question in Experiment 2.

Condition	N	Belief	Reality	Memory
Discriminative				
Correct	42	2.93 (3.89)	1.44 (1.53)	1.90 (1.53)
Incorrect	15	4.25 (3.87)	1.41 (1.83)	1.90 (2.45)
Aspect				
Correct	40	2.58 (3.74)	1.10 (1.07)	2.18 (2.79)
Incorrect	17	4.53 (4.06)	1.44 (1.27)	2.64 (1.98)

Note. The values in parentheses are SDs.

### **3.3 Discussion**

The results replicate Experiment 1 in demonstrating that the response time on the belief question was longer than on the reality and the memory question and that it makes no difference whether the response is correct or incorrect. This is consistent with the prediction from the bias theory.

The difference between Aspect and Discriminative conditions was not significant. Apparently, the Aspect condition is not very effective in helping children to give correct judgments, supporting our suggestion made in the Discussion of Experiment 1 that young children's bias to current reality is rather robust. It is difficult to specify precisely why the replication failed, though it is worth noting that various differences in the stories, the mean age of the sample and the character of the design could have had an impact.

### **4. Experiment 3**

Is it possible that the long response time to the belief question might be due to the time taken to comprehend that question rather than the processing required to resolve the representational conflict between belief and reality? Specifically, perhaps it is more difficult for children to comprehend the belief question than the memory and the reality question (Lewis and Osborne, 1990; Siegal and Beattie, 1991; Wimmer and Hartl, 1991). For example, Siegal and Beattie (1991) reported that the number of children who responded correctly increased when the experimenter asked a less ambiguous question.

In Experiment 3, the response time on true belief as well as false belief questions was measured to examine whether the extended latency is due to comprehension of the question or differences in the calculation of the response associated with representational conflict. The experimenter asked children about the beliefs of two protagonists on the unexpected transfer task. One protagonist held a true belief and another held a false belief. Children were asked a standard belief question (the false belief question), 'Where will X1 (the first protagonist: a person who does not know current reality) look for Y1 (object)?' On the true belief question, they were asked, 'Where will X2 (the second protagonist: a person who knows the current state of reality) look for Y1?' The true belief question has the same structure as the false belief question while the calculation of belief differs because there is no need to resolve representational conflict.

As the wording of the true belief question is the same as the false belief question, the time children needed to comprehend the true belief question would be the same as the time they needed to comprehend the false belief question. However, we predict that calculating a false belief would take longer than calculating a true belief. The false belief question will lead children to consider two possibilities, as the place associated with the protagonist's belief is different from the place associated with the participant's belief: The prior location and the current location of the object. In contrast, the true belief question only allows one possibility: The

place associated with the protagonist's belief is the same as the place associated with the participant's belief and therefore children would not have to select between representations on the true belief question. Conflict between representations is only experienced when children are presented with the false belief question but not when presented with the true belief question (c.f. Leslie and Thaiss, 1992; Roth and Leslie, 1998). This conflict engendered by processing false belief would need to be resolved, and doing so might occupy a certain period of time. Hence, we predict a longer response time in answer to the false belief question than in answer to the true belief question.

If participants did respond more quickly to the true belief question than to the false belief question, then we would have a further opportunity to test the conceptual change account. According to this, children who give an incorrect judgment in a test of false belief effectively treat the task as though it concerned true belief (e.g. Gopnik, 1993; Perner, 1991; Wellman, 1990). If this view is correct, then the response time of a child who answers a false belief question incorrectly should be equivalent to the response time of a child who answers a true belief question correctly.

The second purpose of Experiment 3 was to further investigate effects associated with the Aspect procedure. Specifically, the difference between conditions in terms of response time should be significant on the false belief question, because the Aspect procedure might help children to resolve the representational conflict more speedily. On the true belief question, in contrast, children will not experience representational conflict and so response time should not vary across the Aspect and Discriminative conditions.

The results of Experiment 3, incidentally, also provide an opportunity to examine whether the long response time to a belief question is not so much a reflection of processing time but rather is a sign of hesitation and uncertainty. Ruffman, Garnham, Import and Connolly (2001) found that many false belief failers seemed to have confidence in their incorrect judgment, while false belief passers varied in their level of confidence. The younger members of the sample of passers seemed to have considerably lower confidence than the older members of the sample of passers. Therefore, if a long response time is merely an index of low confidence then we might find that among false belief passers, older children respond more quickly than younger children specifically in response to the belief question. Indeed, we might find that older children are so 'confident' in their correct belief judgment that they respond as rapidly as they do to the memory and reality questions. Experiment 3 included two age groups, thereby allowing scope to test this hypothesis.

## **4.1 Method**

**4.1.1 Participants.** Participants were 57 3-year-olds (mean age 3:06, range 3:01-3:11) and 57 4-year-olds (mean age 4:06, range 4:00-4:11). All children attended a kindergarten of Osaka prefecture in Japan situated in a middle class area.

The numbers of boys and girls aged 3 years were 30 and 27 respectively, while the numbers of boys and girls aged 4 years were 38 and 19 respectively.

**4.1.2 Materials.** Children were presented with the story developed for Experiment 1, except a true belief question was added (see below).

**4.1.3 Procedure.** All children were tested individually, half were in the Aspect condition and the other half in the Discriminative condition. Children in each condition received four questions on the unexpected transfer task: The false belief question, the true belief question, the reality question and the memory question. The order of questions was randomized.

The true belief, the memory, and the reality question were the same for all children while the false belief question differed, depending on condition. Children in the Aspect condition were asked, 'Where will Taro look for his hat?' In contrast, children in the Discriminative condition were asked, 'Where will Taro look for his Plasticine?' On the true belief question, all children were asked, 'Where will Hanako look for the Plasticine?' On the memory question, all children were asked, 'Where did Taro put the Plasticine?' On the reality question, all children were asked, 'Where did Hanako put the Plasticine?'

## 4.2 Results

**4.2.1 Children's Judgments.** Table 4 shows the percentage of children who answered correctly on each question. A Chi-square test indicated that on the false belief question the difference between the Aspect and the Discriminative condition was not significant for older children ( $\chi^2 = 1.25$ ,  $df = 1$ ,  $n = 57$ ) and younger children ( $\chi^2 = 1.93$ ,  $df = 1$ ,  $n = 57$ ). The difference between conditions was not significant on the true belief question (older children,  $\chi^2 = 0.14$ ,  $df = 1$ ,  $n = 57$ ; younger children,  $\chi^2 = 0.04$ ,  $df = 1$ ,  $n = 57$ ), on the memory question (older children,  $\chi^2 = 0.00$ ,  $df = 1$ ,  $n = 57$ ; younger children,  $\chi^2 = 0.00$ ,  $df = 1$ ,  $n = 57$ ) and on the reality question (older children,  $\chi^2 = 0.04$ ,  $df = 1$ ,  $n = 57$ ; younger children,  $\chi^2 = 0.04$ ,  $df = 1$ ,  $n = 57$ ).

**Table 4.** Percentage of children who answered correctly on each question in Experiment 3.

Condition	4 year old Children					3 year old Children				
	N	FBQ	TBQ	RQ	MQ	N	FBQ	TBQ	RQ	MQ
Aspect	28	78.57 (22)	100.00 (28)	100.00 (28)	92.86 (26)	28	57.14 (16)	89.29 (25)	100.00 (28)	71.42 (20)
Discriminative	29	55.17 (16)	93.10 (27)	100.00 (29)	96.55 (28)	29	34.48 (10)	96.55 (28)	93.10 (27)	75.86 (22)

Note. The values in the parentheses are the number of children who answered correctly. FBQ is the false belief question. TBQ is the true belief question. RQ is the reality question. MQ is the memory question.

However, when the two age groups were combined, the difference between the Aspect and the Discriminative condition was significant ( $\chi^2=5.24$ ,  $df=1$ ,  $n=57$ ,  $p<.05$ ) for the false belief question. The difference between conditions was not significant on the true belief ( $\chi^2=0.00$ ,  $df=1$ ,  $n=57$ ), on the memory question ( $\chi^2=0.02$ ,  $df=1$ ,  $n=57$ ) and on the reality question ( $\chi^2=0.17$ ,  $df=1$ ,  $n=76$ ). The results suggest that the Aspect procedure facilitates children to respond correctly on the false belief question although the magnitude of effect is very small.

**4.2.2 Response time.** Table 5 shows the mean response time for each question and in each age group. The response times on each question were analyzed for children who gave a correct or incorrect judgment on the false belief question (response type). The MANOVA of four mixed factors (Age x Condition x Response type x Question) indicates that the main effects of age [ $F(1,96) = 9.22$ ,  $p<.01$ ] and question [ $F(1,96) = 31.70$ ,  $p<.01$ ] and the interaction of condition and age [ $F(1,96) = 5.53$ ,  $p<.05$ ] were significant. None of the other effects approached significance. Paired comparisons using  $t$ -tests to examine the main effect of the question indicate that the response time on the false belief question was significantly longer than the true belief, the memory and the reality question ( $ps<.05$ ), and that the response time on the reality question was shorter than the memory and the true belief question ( $ps<.05$ ) but the response time on the true belief question was not different from the memory question. The multiple  $t$ -test comparisons on the interaction of age and condition indicate that the response time of 3-year-old children was significantly longer than the 4-year-old children on the Aspect condition ( $p<.05$ ) but not on the Discriminative condition.

**Table 5.** Mean response time of children who responded correctly and incorrectly on the false belief question in Experiment 3.

Condition	Correct Response					Incorrect Response				
	N	FBQ	TBQ	RQ	MQ	N	FBQ	TBQ	RQ	MQ
3 year old children										
Aspect	14	9.10 (5.96)	3.36 (2.38)	2.56 (2.06)	2.59 (2.30)	10	6.89 (6.57)	3.29 (3.29)	1.64 (1.80)	4.75 (4.06)
Discriminative	9	6.87 (5.37)	3.17 (1.78)	2.49 (1.84)	4.78 (4.73)	17	6.62 (5.66)	2.69 (2.07)	1.19 (1.05)	3.90 (3.50)
4 year old children										
Aspect	21	3.50 (4.01)	1.86 (1.81)	0.89 (1.06)	1.70 (2.49)	5	3.68 (5.57)	1.92 (1.29)	1.35 (0.89)	1.23 (1.79)
Discriminative	16	6.13 (4.37)	2.54 (2.65)	1.92 (2.68)	2.72 (3.98)	12	7.05 (4.59)	3.33 (1.94)	2.17 (2.52)	3.54 (3.15)

Note. The values in parentheses are the SDs. FBQ is the false belief question. TBQ is the true belief question. RQ is the reality question. MQ is the memory question.

### **4.3 Discussion**

Generally, many of the response times were longer in Experiment 3 than in the previous experiments. It is difficult to place an interpretation on this, other than to suggest that the increased number of questions may have had a generally adverse effect on the speed of response.

As predicted, response time on the false belief question was longer than that on the true belief question. Believers in conceptual change argue that young children, who have not yet acquired the necessary concept, would treat a question about false belief as if it were a question about true belief (e.g. Gopnik, 1993; Perner, 1991; Wellman, 1990). Specifically it is expected from the conceptual change theory that children who were incorrect on the false belief question perform the same processing as children who were correct on the true belief question. The results of Experiment 3 refute this suggestion. Children who were incorrect on the false belief question apparently performed longer processing than children who were correct on the true belief question. Therefore, children discriminated between true and false belief, supposedly an impossible feat among those who lack a concept of belief.

Interestingly, children answered the question about current reality more rapidly than they answered the question about true belief. These two questions required the same answer, and in that context, we need to explain why answers to the true belief question occupied more time. Answers to the reality question merely depended on retrieval, while answers to the true belief question might additionally depend on calculation, albeit calculation that was not protracted by representational conflict. This suggestion must remain tenuous, however, given that response time to the true belief question was not significantly longer than that to the memory question.

The Aspect manipulation should help children to resolve representational conflict. So, it was expected that the response time on the false belief question for the Aspect condition would be shorter than that on the Discriminative condition, and this expectation was supported by the results, although the effect generalized across the range of questions.

There was little evidence to support the suggestion that the long response times indicated low confidence. Although 3-year-olds were a little slower than 4-year-olds, this was confined to the Aspect condition and within the Aspect condition the effect seemed to generalize across the various questions. There was no interaction between age and response type, and so there was nothing to suggest that the variation in speed of response (as a measure of confidence) associated with children's age was mediated by whether they answered the false belief question correctly or incorrectly.

Finally, the results threw up an interaction that is difficult to interpret. Members of the younger age group took longer to respond than members of the older group specifically in the Aspect condition. This effect generalized across questions and therefore it is difficult to understand why the wording of the false belief question should have been relevant to the time taken to answer the control questions specifically in the Aspect condition and specifically among the younger children.

## 5. General Discussion

### 5.1 Response Time as a Clue to How Beliefs are Processed

The main purpose of this research was to examine whether the bias theory gains support from response time data. We proposed that the processing elicited by the belief question involves calculation, and in the case of false belief, this calculation would be protracted by children resolving conflict arising from their knowledge of current reality. In contrast, processing elicited by the reality and the memory question would merely require retrieval of information. It was expected, therefore, that response times on the belief questions would be significantly longer than on the reality and the memory questions. The results of the three experiments consistently support such a prediction. These findings could probably be interpreted in a way that supports variants of the bias theory, including the influential account formulated by Alan Leslie. However, our specific predictions arose from a theory in which beliefs are processed when a query is posed (Peterson and Riggs, 1999), something that is currently beyond the scope of Leslie's theory.

Another focal interest lay in differences in response time between children who gave correct or incorrect judgments of false belief. The main effect and associated interactions of this response time were not significant in any of the experiments. These results raise the possibility that even children who responded incorrectly on the belief question were genuinely trying to calculate a belief response, and were not merely retrieving information (Mitchell and Kikuno, 2000). In other words, these results suggest even if children get the content wrong about another person's belief, they perform the appropriate kind of processing.

Finally, we examined whether the long response time on the belief question was due to difficulty in comprehension rather than the process of calculation. This was examined in Experiment 3 by asking a true belief question that had the same structure as the false belief question. We assumed that both kinds of question required children to calculate a response, but only the false belief question entailed representational conflict and therefore required time-consuming conflict resolution. As predicted, the response time on the false belief question was longer than that on the true belief question. This result suggests that the longer response time on the false belief question was due to the processing involved in conflict resolution rather than difficulty with comprehension.

Children took longer to respond to the true belief question than to the question about current reality, even though both questions required the same response. This might be because an answer to the reality question merely depends on retrieval, while an answer to the true belief question depends on calculation, albeit calculation that is not protracted by having to resolve representational conflict.

### 5.2 The Effect of the Aspect Procedure

We examined whether it was possible to help younger children give correct judgments of false belief by supporting them in selecting the correct content for



belief calculation in the Aspect condition. The results of Experiments 1 and 3 indicate that correct judgments on the Aspect condition were more common than on the Discriminative condition. In Experiment 2, though, there were a slightly higher percentage of correct judgments in the discriminative than in the aspect condition (the contrast was far from significant). These results suggest that the Aspect condition might sometimes help children to select the appropriate content for a judgment of false belief, although the effect was weak and fleeting.

Why didn't the Aspect treatment strongly facilitate correct judgments in response to the false belief question? One possibility is that young children have little potential to process false belief correctly. Perhaps the procedure of the Aspect condition is not strong enough to facilitate performance to a large degree. Children's focus on their own knowledge might be so robust that it is immune to any benefits that could arise from the subtle features of story and question wording in the Aspect condition.

### **5.3 Conclusion**

The results of three experiments were consistent with predictions from the bias theory but offer little support for the conceptual change theory. They suggest that children's judgments of false belief involve a process of calculation. They also suggest that difficulty in acknowledging false belief in 3-year-old children might be caused by limitation in processing that leads to incorrect selection of output content (c.f. Leslie and Thaiss, 1992; Roth and Leslie, 1998) and not basic deficiency in the concept of mind. The results raise the possibility that development in understanding minds depends on the development of the processing mechanism rather than the acquisition of a concept of belief.

*(HK) School of Human Science  
Osaka Shoin Women's University, Japan*

*(PM and FZ) School of Psychology  
University of Nottingham, UK*

## **Appendix 1**

Stories and questions of the unexpected transfer task in Experiment 2

### **Story about Taro:**

A boy whose name is Taro got his Plasticine hat out of the green drawer. He decided to play with his (Plasticine) hat. He was very happy with his hat. Then he moved it to the yellow cupboard. He left the room to play outside. While he was away, his sister whose name is Hanako arrived. She got out the hat. She made the

Plasticine hat into a Plasticine apple and moved it from the yellow cupboard back to the green drawer. Subsequently Taro was returning.

Question:

Belief question:

Aspect condition: 'Where will Taro look for his hat?' Correct answer = 'yellow cupboard'

Discriminative condition: 'Where will Taro look for his Plasticine?' Correct answer = 'yellow cupboard'

Reality question, 'Where is the Plasticine right now?' Correct answer = 'green drawer'

Memory question, 'Where was Taro's hat at very beginning?' Correct answer = 'green drawer'

Story about Rie:

A girl whose name is Rie took her paper from behind the settee. She made it into a paper plane and threw it. She looked at her plane with pleasure. She had to go shopping. She thought that she would play with the paper plane after shopping. She picked it up again and put it away in the box. Rie left the room to go shopping. Her brother whose name is Yoshio arrived. He got the plane out of the box and changed it into a dog. Yoshio moved the paper dog behind the settee. Then Rie returned.

Question:

Belief question:

Aspect condition: 'Where will Rie look for her paper plane?' Correct answer = 'in the box'

Discriminative condition: 'Where will Rie look for her paper?' Correct answer = 'in the box'

Reality question: 'Where is the paper right now?' Correct answer = 'behind the settee'

Memory question: 'Where was Rie's paper in the very beginning?' Correct answer = 'behind the settee'

## References

- Dennett, D.C. 1978: *Brainstorms: philosophical essays on mind and psychology*. Cambridge, MA: MIT Press.
- Flynn, E., O'Malley, C. and Wood, D. 2004: A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, 7, 103–115.
- Fodor, J.A. 1992: A theory of the child's theory of mind. *Cognition*, 44, 283–296.
- Frye, D., Zelazo, P.D. and Palfai, T. 1995: Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gopnik, A. 1993: How we know our minds—the illusion of 1st-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.

- Gopnik, A. and Astington, J.W. 1988: Children's understanding of representational change, and its relation to the understanding of false belief. *Child Development*, 59, 26–37.
- Gordon, A.C.L. and Olson, D.R. 1998: The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, 68, 70–83.
- Harris, P.L. 1991: The work of the imagination. In A. Whiten (ed.), *Natural Theories of Mind*. Oxford: Blackwell, pp. 283–304.
- Keysar, B., Lin, S.H. and Barr, D.J. 2003: Limits on theory of mind use in adults. *Cognition*, 89, 25–41.
- Leslie, A.M. and Thaiss, L. 1992: Domain specificity in conceptual development—neuropsychological evidence from autism. *Cognition*, 43, 225–251.
- Lewis, C. and Osborne, A. 1990: Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact? *Child Development*, 61, 1514–1519.
- Mitchell, P. 1996: *Acquiring a Conception of Mind: A Review of Psychological Research and Theory*. Hove: Psychology Press.
- Mitchell, P. and Kikuno, H. 2000: Belief as construction: inference and processing bias. In P. Mitchell and K.J. Riggs (eds), *Children's Reasoning and the Mind*. Hove: Psychology Press.
- Mitchell, P. and Lacohee, H. 1991: Children's early understanding of false belief. *Cognition*, 39, 107–127.
- Mitchell, P., Robinson, E.J., Isaacs, J.E. and Nye, R.M. 1996: Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition*, 59, 1–21.
- Nisbett, R.E. 2003: *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. New York: Free Press.
- Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Peterson, D.M. and Riggs, K. J. 1999: Adaptive modelling and mindreading. *Mind & Language*, 14, 80–112.
- Premack, D. and Woodruff, G. 1978: Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526.
- Riggs, K.J., Peterson, D.M., Robinson, E.J. and Mitchell, P. 1998: Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13, 73–90.
- Roth, D. and Leslie, A.M. 1998: Solving belief problems: toward a task analysis. *Cognition*, 66, 1–31.
- Ruffman, T., Garnham, W., Import, A. and Connolly, D. 2001: Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, 80, 201–224.
- Saltmarsh, R. and Mitchell, P. 1998: Young children's difficulty acknowledging false belief: realism and deception. *Journal of Experimental Child Psychology*, 69, 3–21.
- Saltmarsh, R., Mitchell, P. and Robinson, E. 1995: Realism and children's early grasp of mental representation—belief-based judgments in the State Change Task. *Cognition*, 57, 297–325.

- Shah, A. and Frith, U. 1983: An islet of ability in autistic-children—a research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 24, 613–620.
- Shah, A. and Frith, U. 1993: Why do autistic individuals show superior performance on the Block Design Task. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 34, 1351–1364.
- Siegal, M. and Beattie, K. 1991: Where to look first for children's knowledge of false beliefs. *Cognition*, 38, 1–12.
- Wellman, H.M. 1990: *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
- Wellman, H.M., Cross, D. and Watson, J. 2001: Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72, 655–684.
- Wimmer, H. and Hartl, M. 1991: Against the Cartesian view on mind—young children's difficulty with own false beliefs. *British Journal of Developmental Psychology*, 9, 125–138.
- Wimmer, H. and Perner, J. 1983: Beliefs about beliefs—representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Zelazo, P.D. 2000: Self-reflection and the development of consciously controlled processing. In P. Mitchell and K.J. Riggs (eds), *Children's Reasoning and the Mind*. Hove: Psychology Press, pp. 169–189.