British Journal of Developmental Psychology (2002), 20, 393–420

www.bps.org.uk

A new false belief test for 36-month-olds

Malinda Carpenter*, Josep Call and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology, Germany

We report two studies that suggest that some 36-month-old (and younger) children understand others' false beliefs. In the false belief conditions, children and two adults (E1 and E2) watched as an object was put into a container. E1 left the room, and E2 switched that object with another. E1 returned, expressed her desire for the object, and struggled to open the container (without succeeding). She spied both objects across the room and said to the child. 'Oh, there it is. Can you get it for me?' In other conditions, the object was not switched or E1 witnessed the switch. Other variations included using a novel word for the object and removing the 'pull of the real'. Measures of children's latency and uncertainty were taken to determine whether those who were correct on the false belief tasks were guessing (luckily) or truly understanding. Results showed that between one-third and almost two-thirds of children took account of the adult's false belief when deciding which object the adult was requesting. We compare this task with other implicit and explicit tasks. We also conclude that certain task demands (e.g. the pull of the real) in traditional theory of mind assessments are unnecessary.

In a meta-analysis of more than 100 studies of children's performance on traditional false belief tasks, Wellman, Cross, and Watson (2001) found a reliable age trend. Whereas children age 4;0 and older perform at above chance levels as a group on these tasks, children age 3;5 and younger perform at below chance levels as a group. These younger children make the classic false belief error of predicting that, for example, a story character will look for a toy where it now is hidden (after it was moved during her absence) instead of where they originally saw it hidden.

There are at least two reasons why children younger than 3;5 might consistently fail false belief tests. One is that they simply do not understand that other people can have beliefs that do not match with reality as they themselves know it. Another is that, at least by a certain age, they do have this understanding but that there is something about the traditional false belief tests that lead them to give an incorrect response. There is

* Requests for reprints should be addressed to Malinda Carpenter, Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103 Leipzig, Germany (e-mail: carpenter@eva.mpg.de).

much recent evidence that supports the latter possibility.

There have been a number of lines of research that demonstrate that 3-year-old children show some early evidence of false belief understanding. First, in their natural interactions with others, these children often talk and behave in ways that seemingly evidence their understanding of the thoughts and beliefs of others—even when those thoughts and beliefs differ from the child's own or from reality (Bartsch & Wellman, 1995; Dunn, 1988). Second, some experiments using measures other than verbal responses have also found some evidence in this same direction, for example, acts of deception (Chandler, Fritz, & Hala, 1989; Polak & Harris, 1999; see also studies using pretence, e.g. Cassidy, 1998).

The study that has elicited the best performance from the youngest children was that of Clements and Perner (1994) (and see Garnham & Ruffman, 2001 for a replication). In that study, children were presented with a standard change of location task in which an object was moved from one box to another during a story protagonist's absence. However, in addition to the traditional action prediction question ('which box will he open first?'), Clements and Perner also measured to which location children looked in anticipation of the protagonist's return. They found that although most children could not answer the action prediction question correctly until about age 4 years, beginning at age 2;11, children looked more often in anticipation to the correct location—the location where the character had originally seen the object placed instead of the object's current location. This indicated to Clements and Perner that children as young as 2;11 have an 'implicit' understanding of false belief. Importantly, in that study, there was also a true belief control condition in which the location of the object was switched in full view of the protagonist. In this condition, almost all children looked correctly to the current location of the object.

There are several reasons why children might show some implicit or naturalistic understanding of others' false beliefs but still fail the traditional false belief tests. It is well known that children's performance in false belief tasks is fairly sensitive to seemingly minor changes in extraneous task demands. For example, in the traditional false belief tasks, children are asked to predict what someone will do. When children are asked instead what a person wants (a character says they want 'the one that's in the red drawer', although the object they think is in the drawer has actually been moved, and another is in its place), they find the task much easier (Robinson & Mitchell, 1992). Also, not unexpectedly, another powerful variable is the 'pull of the real'. When children are told a story in which the location of an object is switched while the story character is out of the room, but then the desired object is eaten before she returns (so that the desired object has no real location at the time of choice), children again find the task much easier (Koos, Gergeley, Csibra, & Biro, 1997; Wellman *et al.*, 2001; Wimmer & Perner, 1983).

Importantly, neither of these task modifications—having the child retrieve an asked-for object (rather than predict what someone will do) and removing the pull of the real—changes the basic problem facing the child. The story character still has a false belief (they think the desired toy is still where they saw it placed), and the child must take that false belief into account in order to respond correctly.

In the current studies, we introduce a new paradigm for assessing children's understanding of false belief that includes both of these modifications. We do not require children to predict what a protagonist will do, but at the same time, we require a more active behavioural response than just looking to the correct location. Children participated in a real interaction and then had to choose which of two objects an adult wanted. The paradigm was adapted from the word learning studies of Tomasello and colleagues (reviewed in Tomasello, 2001; see especially Akhtar & Tomasello, 1996). In the first study, the child and an experimenter put a novel object in a container with an adult watching, and then, after the adult left the room, switched it for another novel object. When the adult returned, she gestured towards the container and tried to open it, saying she wanted to get her 'dax' out. When she could not open it, she gave up. Subsequently, in a retrieval test with both objects across the room, the adult asked the child to go get the 'dax'. If children understood that the adult had a false belief about which object was in the box, they should know that the object the adult was intending to name was the one that she had originally seen put there, not the one that was there now. If children did not understand the adult's false belief, then they should think that the adult was naming the object in the box currently. In the second study, the same general procedure was followed, but the pull of the real was removed. When the adult returned, the object had already been removed from the container, so the adult was gesturing towards an empty container instead of one that contained the distracter object.

We also addressed a methodological issue in the current studies. Typically, children's performance on false belief tasks is compared with 50% chance levels. There are two problems with this approach. First, at-chance responding for the group as a whole may be produced by either, at the extremes: (1) all children randomly guessing or (2) all children confidently choosing an answer with half being right and half being wrong. One approach to differentiating children who are guessing from children who are making a considered choice based on an understanding of the task is to find some independent measure that correlates with children's real choices versus guesses. A good candidate for this independent measure is response latency, as it is reasonable to assume that children will choose relatively quickly when they are responding with genuine understanding but they will hesitate or pause when they are not sure of themselves. Similarly, children sometimes express overtly to experimenters that they are unsure of the choice they are making by doing such things as looking at them quizzically or asking them 'this one?' Neither of these measures is perfect, of course, because children may act more quickly or slowly for a number of different reasons, and even children who are fairly certain of their choice may sometimes check with adults nevertheless; ultimately, such measures need validating. But, again, the main point is that these measures may potentially provide additional information to help in evaluating the performance of individual children.

Second, the comparison with 50% chance can be misleading. For instance, if children do not understand others' false beliefs, they will treat the adult as having a true belief. If this is the case, they should respond identically (i.e. in the current study, they should choose the same object) in both the true belief and the false belief conditions. In addition, any such children who failed the true belief test would pass the false belief test for the wrong reason—that is, not because they understand the adult's false belief but simply because they are using some irrelevant strategy that happens to be consistent with passing the false belief test. Another reason the comparison with 50% chance can be misleading is that when there is a pull of the real, children may be drawn to the reality answer, even when they are guessing, so the true chance value in this case might not be 50% Then, comparing performance with that in the true belief condition is a better measure. Thus, we believe that in many cases, the most appropriate comparison for false belief conditions is with children's performance in a true belief condition, not with 50% chance.

To address these methodological issues, in both of the current studies, we compare latency and uncertainty measures in the false belief condition with a condition in which children are forced to guess which object to choose. In Study 1, we also compare children's performance in the false belief condition with what we would expect given their performance in a true belief condition.

STUDY 1

In addition to a false belief condition, in which the object in the container was switched while the adult was out of the room, in this study, there were two other conditions. In a true belief condition, the object in the container was not switched while the adult was away—the first object was simply taken out and put back in the container. The addition of a true belief condition was important in this study for the reasons discussed above and because we were using a new procedure, and requiring children to learn a novel word, so we needed to know what percentage of children had trouble with the basic procedure, regardless of their understanding of others' false belief. We thus compared children's tendency to choose the correct object in the false belief condition (i.e. the object outside the container when the adult returned) with their tendency to choose the same object (the incorrect one in this case) in the true belief condition.

We also included a third condition, in which children were forced to guess which of the two objects the adult wanted. In this control condition, which also tested for methodological artifacts in the false belief test, the same procedure as the false belief test was followed except that instead of using the novel word while trying to open the container, the adult used the novel word for the first time during the retrieval test. Thus, at the time of the test, children in the control condition could not know which object the adult wanted.

We predicted that children would do better in this new false belief test than we would expect given the below-chance performance of other children at the same age on traditional false belief tests. We compared latency and uncertainty measures in the false belief and control conditions to determine whether children were guessing or deliberately choosing (correctly or incorrectly) in the false belief condition.

Method

Participants

Participants were 24 American children (13 girls and 11 boys) who were within 4 months from their third birthday (mean age = 35.2 months; range = 2;8;2 to 3;3;18). Children were recruited from daycare centres or from music teachers and were all native English speakers.

Materials and design

Six objects were chosen on the basis that they were considered likely to be novel, nameless, and equally interesting to children. Objects were: a metal counter, interconnected PVC tubes, metal tongs, a modified tape measure, a tube with a string bracelet through it, and a key chain with rubber strands hanging from it. An additional novel, nameless object (a top) was used as a substitute if the child produced a name for one of the others before the game began (this happened only once). For hiding objects, we used three containers, one in each condition (a box, a bucket, and a small chest of drawers).

Each child participated in three conditions, counterbalanced for order across children. Each condition used two randomly selected, novel, nameless objects, one randomly designated as the target and the other randomly designated as the non-target. One of three novel words (*toma, dax, or gazzer*) was randomly selected and used to refer to the target object. The three conditions were as follows:

True Belief: An adult watched as the target object was placed in a container. She then left the room. The target object was taken out of the container and put back in, and the adult returned and announced her intention to take the 'gazzer' (or 'toma' or 'dax') from the container.

False Belief: An adult watched as the target object was placed in a container. She then left the room. The object was 'sneakily' switched with another object, and the adult returned and announced her intention to take the 'toma' (or 'dax' or 'gazzer') from the container.

Control: This condition was exactly like the False Belief condition, except that on returning the adult simply announced her intention to take the 'toy' from the container (i.e. no new word was used until the retrieval test, see below). This condition served two purposes. First, it was a control for methodological artifacts: if children performed above chance in this condition, and did well in the False Belief condition, then we could assume that some procedural factor other than the adult's false belief (e.g. different highlighting or handling of objects) could explain children's choice of the correct object in the False Belief condition. Second, the first time children in the Control condition heard the novel word was in the retrieval test (and so they had no way of knowing which object the adult was referring to). This thus represented a situation in which guessing (50% chance responding) was expected.

Procedure

A total of 21 children were tested individually in a quiet room in their daycare centre, three children were tested in their home. Each child sat on the floor with two experimenters, E1 and E2. After a brief warm-up period with unrelated toys, E1 took two novel objects from a bag, placed them in front of the child, and encouraged her to play with them. Both Es used only pronouns (*it, this one*) to refer to the objects, never object labels. The object that the child chose to touch first was noted for preference analyses. The child was allowed to play with the objects for approximately 1 min, during which time, the Es pointed out interesting things that could be done with each object, and made sure that the child handled each object approximately equally. Then, E1 suggested playing a game with the objects. She brought out a container and, while the child and E2 watched, she put the target object inside the container and the non-target object to the side of the container (on the floor), describing her actions all the while ('Look, I'm putting this one in here and this one over here'), and she closed the container. E2 then took the objects out of the container and from the floor and placed them back in front of the child.

This game of putting the target object in the container and the non-target object to the side of the container was repeated three more times. Each time, except for the last time, E2 retrieved the objects and gave them back to the child. In the last repetition, after the objects were in their respective places (i.e. the target object in the closed container and the non-target object on the floor), E2 excused herself (to go to the bathroom, speak to a teacher, or get a drink or a tissue) and left, closing the door behind her.

While E2 was gone, E1 retrieved the objects and gave them back to the child, and then one of two events took place, depending on the experimental condition. In the True Belief condition, E1 simply played the same game again (she put the target object back into the container) with the exception that she now put the non-target object into her bag, out of sight ('Let's put this one away'), before closing the container. In the False Belief and Control conditions, however, E1 told the child that they were going to play a trick on E2. Instead of putting the target object into the container again, she put the non-target object into the container, remarking that she was 'switching them'. Then, she put the target object into her bag, out of sight ('Let's *hide* this one'), and closed the container. All this was done in a 'sneaky' manner, with E1 whispering, looking often at the door to see if E2 was about to return, and asking children to not tell E2 what they had done.

Then, in all conditions, E1 went to the door to call E2 back. Note that neither of the objects was visually present at this time. When E2 entered the room, she announced that she wanted to get an object, with the way she did this differing in the different conditions. In the True Belief and False Belief conditions, she used a novel word (i.e. *dax*, *toma*, or *gazzer*) eight times: three times while walking over to the container, and five times while kneeling beside and holding on to the container and alternating gaze between the container and the child or E1 (e.g. [upon entering] 'Hey, let's get the dax. I'm going to get the dax. I'll get the dax and then we can play with it'. [At the container:] 'Ok, I'm going to get the dax now. I'm going to get the dax out of here'. [To the child:] 'Should we get the dax?' [To E1:] 'Can I get the dax? Ok, I'm going to get the dax'). In the Control condition, the same procedure was followed, but instead of using a novel word, E2 used the word *toy* in its place.

E2 then tried unsuccessfully to open the container. After declaring it 'stuck', she suggested that they play with another toy instead. E2 and the child went several metres away and played with the other toy. Meanwhile, E1 put both the target and non-target objects on a nearby chair (on random sides of the chair). For the retrieval test, E2 then looked around the room for something else to play with. She pointed to the chair, without pointing or looking specifically to one of the two objects, and said to the child (using the appropriate novel word): 'Oh, look, there's the ______! Can you go get the ______ and we'll play with it over here.' The child then went to the chair and retrieved an object. If she tried to take both objects, E2 or E1 told her to take 'just the ______'. E2 and the child then played with the chosen object (no feedback was given on whether it was the correct one or not) for a few moments. One could argue that because we gave no feedback, children could assume that their choice was correct. This might make them stick with the same strategy in subsequent trials. However, if children stuck with the same strategy, their performance would be lower because in order to pass both the True Belief and False Belief conditions, children had to switch strategies—choose the object inside or outside the container—depending on the condition.

For each child, each condition had a different pair of objects, a different container, a different novel word, and a different toy. Table 1 presents a comparison of the procedures in the three conditions.

Table 1. Summary of the procedural differences across conditions in Study 1

True Belief	E1: puts A in container and B on floor E2: leaves	E1: (i) takes A out, (ii) puts A in, (iii) puts B away	E2: returns, looks to container, and says, *'I'm going to get the …*	[*dax*].'  E2	E2: (i) tries to open container. (ii) During play with other toys, she looks around and says: *'Oh look, there's the …*	[*dax*] …	*Can you get it for me?'*
False Belief		E1: (i) takes A out, (ii) puts B in, (iii) hides A		[*toma*].'  E2		[*toma*] …	
Control				*toy*.'  E2		[*gazzer*] …	

Coding and reliability

All trials were videotaped. Children's choices in the retrieval test for each condition were recorded live on code sheets by E1 and E2 in collaboration, and then checked by an independent observer on the videotapes. Videotapes were examined to obtain measures of children's latency to choose an object and their uncertainty about their choice of object in the retrieval test. The latency (in seconds) to choose an object was defined as the amount of time from when E2 finished her verbal request to get the [dax] to when children first grasped an object and turned away from the chair. Uncertainty (yes/no) was coded if children looked questioningly to one of the Es with their choice, asked 'This one?' (or something similar) during their choice, or refused to choose. Reliability was assessed for each measure by having another coder recode five randomly selected children (21%) while blind to the experimental condition. For choice in the retrieval test, initial preference, and uncertainty, 100% agreement was achieved. For latency, a Pearson's product-moment correlation between the latencies identified by each coder was .98, $p < .001$, and a *t*-test showed no significant difference between the two sets of scores ($t = .56$, n.s.; the difference in means was 0.46 s).

Results

We analysed the effect of the order of presentation of each condition for the following variables: performance, latency to respond, and signs of uncertainty. Owing to the small samples involved in the performance and uncertainty variables, we collapsed the second and third orders and compared them with the first order for each condition. The order in which children experienced the three experimental conditions had no effect on their performance (Fisher tests: $p > 0.36$ in all cases), latency to respond ($F(2,20) < 2.81$, $p > 0.08$, in all cases), or signs of uncertainty (Fisher tests: $p > 0.09$ in all cases). Consequently, order is not considered in any further analyses.

Main analyses

Figure 1 presents results for the retrieval test. Depicted are the percentage of children in the True Belief and False Belief conditions who chose the object that was inside versus outside the container during E2's language model. In the True Belief condition, the majority of children (83%) correctly selected the object that was inside the container during E2's language model, thus replicating the results of Akhtar and Tomasello's (1996) study in which children also had to recall which object was in a container when they heard a new piece of language. To assess children's false belief understanding, we compared what children did in the False Belief condition with what they did in the True Belief condition as a baseline. We thus compared children's performance in the False Belief condition not with chance (50%) but rather with what they would do if they assumed that the adult was using the novel word for the object inside the container (as in the True Belief condition). The result was that more children in the False Belief condition correctly chose the object outside the container than would be expected based on the True Belief baseline (50% observed versus 17% expected; Binomial test; $p < .001$), and conversely, fewer children incorrectly chose the object inside the container than would be expected based on the True Belief baseline.

As another approach to the data, we also assessed individual children's patterns of

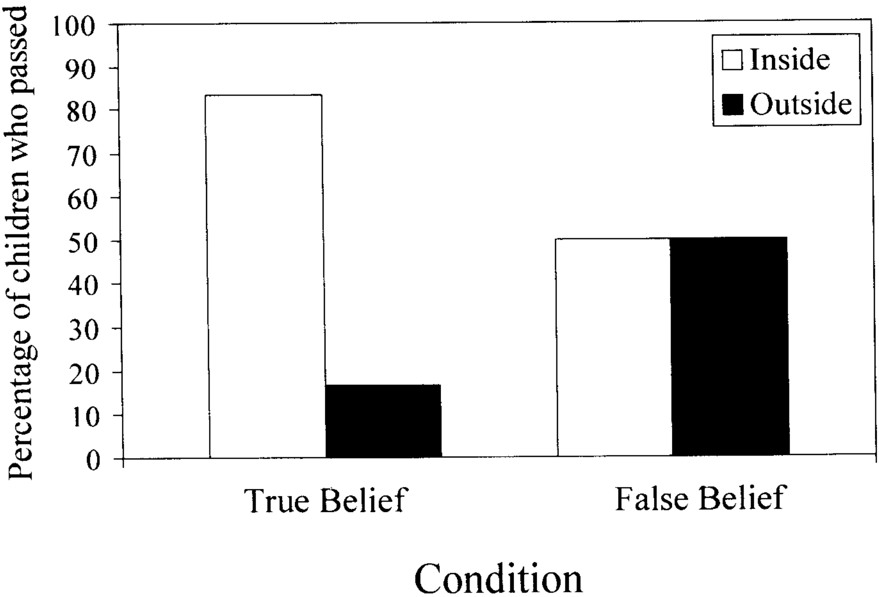


Figure 1. Percentage of children who chose the object that was inside versus outside the container during E2's language model in the True Belief and False Belief conditions' retrieval tests.

responding across the True Belief and False Belief conditions. Table 2 presents the types of strategies used by each of the 24 children across these two conditions. Overall, 12 children chose the object that was inside the container during the language model in both conditions, and four children chose the object that was outside the container during the language model in both conditions; these children thus apparently did not understand E2's false belief. However, eight children (33%) switched strategies depending on the condition. All eight did so in the same way, that is, they correctly chose the object inside the container in the True Belief condition and correctly chose the object outside the container in the False Belief condition. No child showed the reverse pattern (McNemar test: $p < .01$).

These eight children therefore seemed to understand the adult's referential intention in both versions of the finding game, that is, regardless of whether the adult had a true belief or a false belief about which object was currently in the container. If we use the true belief task as a way to screen out children who are confused by the task in general, not paying attention, or not cooperating, then 8 out of 20 children (40%) showed this understanding.

The Control condition was designed to determine whether some methodological artifact was allowing children to choose correctly in the False Belief condition. This was not the case. As expected, children in the Control condition chose randomly: almost equal numbers chose the object that was inside and the object that was outside the container (10 and 14, respectively; Binomial test: n.s.). Children's strategy in the Control condition was not the same as that in the False Belief condition: children who chose the object inside the container in the False Belief condition did not consistently choose the object inside the container in the Control condition, for example, McNemar test: n.s.

There were no significant age differences between children who chose correctly and

Table 2. Individual strategies: number of children who chose the object that was inside versus outside the container during E2's language model in the True Belief and False Belief conditions' retrieval tests

	True Belief	
False Belief	Inside	Outside
Inside	12	0
Outside	8	4

those who chose incorrectly in the retrieval test in any condition (all *t*-tests: n.s.). Half of the eight children who passed both the True Belief and False Belief tests were younger than 2;11. Children did not choose the same object during the preference and retrieval tests above chance in any of the three conditions (Binomial tests: n.s.).

Latency and uncertainty

It must be noted that roughly equal numbers (50–58%) of children chose the outside object in the Control and False Belief conditions. In the Control condition, this performance was as expected because children had no information about the new word's referent when asked to 'bring me the [dax]' in the retrieval test—and so they had to guess. In the case of the False Belief condition, however, it is not likely that children were guessing. We base this conclusion not only on children's use of different strategies in the False Belief and Control conditions, but also on the way children made their choice, that is, *how* they responded to the adult's request in the retrieval tests for the Control and False Belief conditions. We predicted that children would show more signs of hesitation and uncertainty in the Control than the False Belief condition because in the Control condition they had no information and so had to guess. In contrast, in the False Belief condition, children had a hypothesis (either correct or incorrect), and so they should have responded without hesitation. The results of two analyses aimed at this prediction were as follows. First, children took more than twice as long to choose in the Control condition (10.8 s) as in the False Belief condition (5.2 s; $t = 1.96$, d.f. = 22, $p < .05$ one-tailed). Second, more than four times as many children showed signs of uncertainty in the Control condition (9 children) as in the False Belief condition (2 children; McNemar test $p < .05$ one-tailed).

It is also important to note that in the False Belief condition children who were correct did not differ from children who were incorrect on latency to choose ($t = 1.61$, $df = 22$, n.s.) or signs of uncertainty (Fisher test: n.s.). This is important because it shows that even children who chose incorrectly in the False Belief condition were not just guessing, but rather they had an idea (albeit an incorrect idea) about which object the adult intended to indicate with the novel word.

Comparison with a previous study

It is also of interest to compare the findings of the current studies with those of an influential study that used an implicit measure of false belief understanding. Figure 2 presents children's performance in the False Belief condition in Study 1, along with the implicit (looking) and traditional (verbal) False Belief measures from Clements and

Perner (1994). All scores are adjusted for children's performance in the True Belief task; that is, at each age we plotted the number of correct responses in the False Belief task divided by the number of correct responses in the True Belief task. We believe that this adjusted false belief score is the best assessment of children's false belief understanding at the different ages, because it represents the proportion of children who passed the False Belief task out of a population of children who knew enough about how the task worked to be correct on the True Belief task. In some sense, our measure filters out some of the 'noise' due to some children's inability to deal with basic task demands and complex discourse with an adult. What can be seen in Fig. 2 is that the results of the current study compare favourably with those of Clements and Perner (1994), who used a simple looking (expectancy) measure. This is notable because our tasks required children to actually make a decision and retrieve an object for the adult—which would seem to have more behavioural complexity than simply looking.

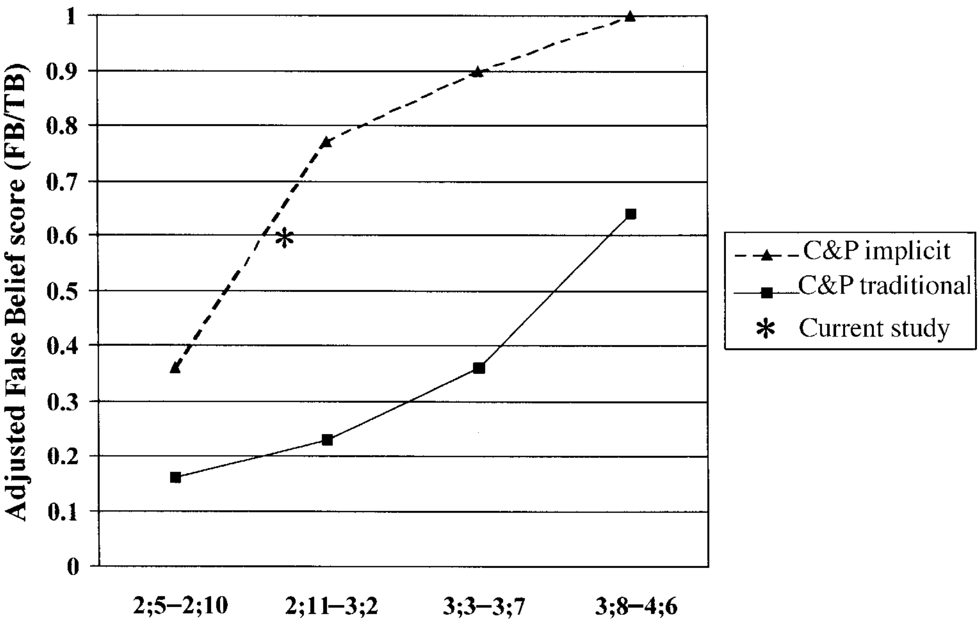


Figure 2. Adjusted false belief scores—i.e. false belief (FB) / true belief (TB)—for children in the Clements and Perner (1994) (C&P) study and the current Study 1. The age range in the current study was 2;8 to 3;4.

Discussion

In this study, we presented children aged 2;8 to 3;4 with two situations in which an adult used a novel word for one of two novel objects, neither of which was visible at the time. In one situation, this adult had a true belief about the location of the object she was referring to, whereas in the other situation, she had a false belief. Children had to figure out which object she was naming. One third of the children tested showed evidence of understanding the adult's false belief by choosing the correct object in the retrieval tests for both these conditions. If we exclude the four children who did not pass the true belief test—under the assumption that they did not understand the general procedure—then 40% of the children chose correctly in both conditions. Children in the false belief condition did not appear to be guessing randomly or uncertain about their choice; they took less time to choose an object and showed fewer signs of uncertainty in the false belief condition than in a control condition in which they had insufficient information and so were forced to guess. Children were relatively quick and confident in the false belief condition irrespective of whether they made the correct choice.

This study thus suggests that some 3-year-olds (up to 40%) can understand other peoples' false beliefs in at least some situations. This group of children in our study thus passed our false belief test a full year earlier than we would expect them to pass traditional false belief tests. That is, based on Wellman *et al.*'s (2001) meta-analysis, we would expect children at this age to perform at below-chance levels as a group on the traditional tests. However, children in our study performed at chance levels as a group if one uses 50%as chance (as in most traditional analyses) or at above-expected levels as a group if one uses children's performance in the true belief condition to calculate baseline expectancy.

The task we gave children was therefore easier than traditional false belief tests. One likely reason for this is that children in this task were not required to answer explicit questions nor to predict what a story character would do next (see the General Discussion for more on this point). Another possible reason is that our procedure involved learning new words. Although, at first glance, it may seem counterintuitive that this would facilitate children's performance, learning new words is something children do several times a day, every day, and in each case, they must in some sense assess the adult's referential (communicative) intentions (Tomasello, 2001). It is thus possible that our word-learning task may have put children in the mindset of 'mindreading' and that this helped them achieve their full potential on this task.

However, there are several limitations to what we can conclude about children's false belief understanding from this task. In terms of percentage correct, children did no better in the false belief condition than in the guessing control condition. We also have no independent verification that what we measured in the latency and uncertainty analyses was really uncertainty. That is, if we had found the opposite result—that children took longer to choose and showed more signs of uncertainity in the false belief condition than the control condition—we could have argued that when children guess they do so quickly and do not have to take the time to consider alternative possibilities as in the more cognitively complex false belief condition. There is also a possible alternate explanation for our findings. For example, in the false belief and control conditions, E1 switched the objects 'sneakily'. Perhaps, for some reason, this caused children to choose the object outside the container more than they would normally have. We also did not compare directly children's performance in the current study with what these children would have done on the traditional test (maybe these children were special in some way and would have passed the traditional test too). We also need to test our word-learning hypothesis directly by comparing children's performance on versions of the same task that use novel words or no novel words. Finally, during her language model, E2 was pointing to, and looking in the direction of, the object that was inside the container (although she could not see that object). We suspect that this made the children's task very difficult—they had to overcome a strong 'pull of the real' (i.e. the object E2 was oriented to) and choose the object that had been removed from the scene. We therefore conducted a second study, in which we attempted to address these limitations and difficulties.

STUDY 2

Study 2 employed the same basic procedure as Study 1, but there were several important improvements. First, we included several false belief trials in order to have a more sensitive measure of children's performance than the single false belief trial in Study 1. Second, we included a validation procedure for the latency and uncertainty measures by presenting children with situations in which they either definitely knew what E2 wanted (because she asked for it by name) or they definitely did not know what E2 wanted (because she was vague about it); these were called the Knowing Baseline and Guessing Baseline conditions. Third, we removed E1's 'sneaky' behaviour. Fourth, we tested the same children on traditional false belief tasks and compared their performance on these with their performance in our new tasks. Finally, we directly tested our word learning hypothesis by having two versions of the task—one in which E2 used a novel word to refer to the object she wanted (as in Study 1) and one in which she used a familiar, generic word (i.e. 'toy').

In addition, to make the task easier for children, we removed the 'pull of the real' by emptying the container before E2 returned and tried to open it. E2 thus believed falsely that a particular toy was inside, when in fact both the target and non-target objects were across the room (and the child knew this).

We included two versions of the false belief condition so that children were required to choose a different object in each type of condition in order to be correct in both. After removing the pull of the real, a true belief condition was no longer possible because when the adult returned and gestured to the empty container, she always had a false belief that there was an object in there. Unlike Study 1, this is not a problem: when there is no pull of the real, and children are truly guessing, chance should be 50% The general procedure was as follows: in one pair of false belief conditions (one with a novel word and one with a generic word), the adult watched as object A was put in the container, and continued watching as A was removed and B was put in. Then, she left the room, and B was taken out of the container. This pair of conditions was therefore very similar to true belief conditions in previous studies in which the protagonist sees the switch of objects (or locations, as in most other tests), except that in this case, when she returned, the adult had a false belief about the presence or absence of the particular object. In the other pair of false belief conditions, the adult did not see the switch of object A to object B; on her return, she thus should believe that object A was in the container, even though B was the last object in there, and there was in reality no object in the container.

We predicted that children would do better in these new false belief tasks than in the false belief task from Study 1 because the 'pull of the real' was removed, and also better in this task than in the traditional false belief tests (for many reasons). We were not sure what to expect in the comparison between the Novel Word and Generic Word conditions, but one possibility was that the extra effort needed to try to discern the adult's communicative intentions when she used a novel word would somehow enhance children's attention to the adult's mental states. We also predicted that children would show longer latencies and more signs of uncertainty in the Guessing Baseline condition than in the Knowing Baseline condition and, more importantly, that these measures of uncertainity in the false belief conditions would pattern with the Knowing Baseline condition—because we hypothesized that in the false belief conditions, children mostly were not guessing but rather had a definite idea (even if wrong) about which object E2 wanted. Importantly, if these predictions were borne out, we could then be more confident in our interpretation of the latency and uncertainty measures in Study 1.

Method

Participants

Participants were 32 German children (16 girls and 16 boys) who were within approximately 2 months of their third birthday (mean age = 36.2 months; range = 2;9;28 to 3;2;17). Children were recruited from their daycare centres in a medium-sized metropolitan area and were all native German speakers.

Materials and design

For the four experimental conditions and three of the baseline conditions, we used 14 novel, nameless objects, chosen to be equally interesting to children (and similar to those in Study 1). There were two additional novel, nameless objects that could be used as substitutes if the child produced a name for one of the 14 before the game began (this happened only five times). For hiding the objects in the four experimental conditions, we used four containers, one in each of the experimental conditions (two different boxes, a small chest of drawers, and a bucket). For the Knowing, Familiar Word Baseline condition, we used two highly familiar objects with well-known names (a toy frog and toy elephant). The warm-up period used six familiar objects and a common box.

Each child participated in all four experimental conditions, counterbalanced for order across children. There were 24 different orders, and six of these randomly selected orders were repeated once or twice. Each condition was first for eight children. Yoked with each of these experimental conditions (and immediately following it in the procedure) was a baseline condition for measuring uncertainty and latency (see below). Each experimental condition used two randomly selected, novel, nameless objects, one randomly designated as the target and the other randomly designated as the non-target. For two of the experimental conditions and one of the baseline conditions, one of three novel German words (*Faudi, Bokel, or Lapus*) was randomly selected and used to refer to the target object; in all but one of the other conditions, the target was designated by the familiar but generic word for 'toy' (*Spielzeug*). The four experimental conditions were as follows:

False Belief B, Novel Word: An adult watched as object A (the non-target object) was placed in a container. Then, she watched as object A was taken out of the container and replaced with object B (the target object). The adult then left the room. Object B was taken out of the container and placed across the room (thus, the container was now empty). When the adult returned, she announced her intention to take the 'Faudi' (or 'Bokel' or 'Lapus') from the container.

False Belief B, Generic Word: This condition used the same procedure as the False

Belief B Novel Word condition with the exception that when the adult returned, she announced her intention to take the 'toy' (i.e. 'Spielzeug') from the container—she did not use a novel word.

False Belief A, Novel Word: An adult watched as object A (the target object) was placed into a container. She then left the room. Then object A was replaced with object B (the non-target object), in a normal, non-sneaky manner. Object B was then taken out of the container and placed across the room (thus, the container was now empty). When the adult returned, she announced her intention to take the 'Bokel' (or 'Lapus' or 'Faudi') from the container.

False Belief A, Generic Word: This condition used the same procedure as the False Belief A Novel Word condition with the exception that when the adult returned, she announced her intention to take the 'toy' (i.e. 'Spielzeug') from the container—she did not use a novel word.

In addition, there were two Knowing baseline conditions (designed to assess latency and signs of uncertainty when the child was certain of what the adult wanted) and two Guessing baseline conditions (designed to assess latency and uncertainty when the child had no idea what the adult wanted). The baseline conditions (in the same order as their corresponding experimental conditions) were as follows:

Knowing, Familiar Word Baseline: In this condition, the adult indicated two familiar objects on a blanket (a toy frog and elephant) and asked the child to 'get the frog'.

Knowing, Designated Generic Word Baseline: In this condition, the adult indicated two novel objects (one on a chair and one on a box, both on the blanket) and asked the child to get 'the toy on the chair'.

Guessing, Novel Word Baseline: In this condition, the adult indicated two novel objects (both lying on the blanket) and asked the child to get the 'Lapus' (or 'Faudi' or 'Bokel').

Guessing, Generic Word Baseline: In this condition, the adult indicated two novel objects (both lying on the blanket) and asked the child to get 'the toy I want'.

In addition, we administered the traditional, verbal versions of the Sally-Anne (change of location) and Smarties (deceptive box) false belief tasks, using the exact procedure and wording (translated into German) as reported in the original papers (Baron-Cohen, Leslie, & Frith, 1985, for Sally-Anne; Perner, Frith, Leslie, & Leekam, 1989, for Smarties). These were presented in counterbalanced order, at the beginning of the session for half the children and at the end for the other half.

Procedure

Children were tested individually in a quiet room in their daycare centre. The child sat on the floor with two female experimenters, E1 and E2. A small blanket was placed on the floor 1.9 m behind the child. First, there was a warm-up game to familiarize children with the procedure of going to get only one of two toys from the blanket. E1 put one of two familiar toys (e.g. a dog and a bear) in a box, closed the box, waited a few seconds, and then took the toy out. She then put both toys on the blanket (0.7 m apart and equidistant from the child) and E2 asked the child to go get 'only the one that was in the box'. Children were corrected if they tried to bring both toys. This game was played a total of three times.

After the warm-up game, E1 took two novel objects from a bag, placed them in front of the child, and encouraged them to play with them. Both Es used only pronouns (the German equivalents of *it* and *this one*) to refer to the objects, never object labels. The object that the child chose to touch first was noted for preference analyses. The child was allowed to play with the objects for approximately 1 min, during which time, the Es pointed out interesting things that could be done with each object, and made sure that the child handled each object approximately equally. Then, E1 suggested playing a game with the objects. She put both objects in her lap, covered them with a cloth bag, and brought out a container. While the child and E2 watched, E1 took out one of the two objects, put it inside the container, tipped the container so that both the child and E2 could see the object inside, and then closed the container.

The ordering of the next two events—switching the objects and E2 leaving the room—was determined by the experimental condition. In both False Belief B conditions, E2 stayed in the room while E1 took the first object out of the container and put it back in her lap, under the bag. E1 then put the other object in the container, tipped the container so the child and E2 could see the object inside, and then closed the container. Then, E2 excused herself (to go to the bathroom, speak to a teacher, or get a drink or a tissue) and left, closing the door behind her. Thus, in these two conditions, the last object E2 saw in the container was the second object. In both False Belief A conditions, E2 watched E1 put the first object into the container, but then she left the room before E1 took that object out of the container (i.e. while the container was still closed, with the first object inside). E1 then switched the objects as in the other conditions, simply saying to the child (in E2's absence) 'OK, let's put this one in there now'. Thus, in these two conditions, the last object E2 saw in the container was the first object.

Then, in all conditions, while E2 was still gone, E1 took the second object out of the container and closed it. The container was thus empty at this point and for the rest of the trial. She then placed both objects (in random order) on the blanket behind the child and went to the door to call E2 back. When E2 entered the room, she looked directly at the container (completely ignoring the toys on the blanket) and announced that she wanted to get an object, with the way she did this differing in the different conditions. In both Novel Word conditions, she used a novel word; in both Generic Word conditions, she used the word 'toy'. In all conditions, she said this word eight times: three times while walking over to the container, and five times while kneeling beside and holding on to the container, alternating gaze between the container and the child or E1 (as in Study 1). E2 then tried unsuccessfully to open the container. She declared it 'stuck' and then casually looked around the room. For the retrieval test, she then turned the child in the direction of the two toys on the blanket (i.e. the child's back was to E2 so that E2 could give no inadvertent cues) and said to the child (using either one of the novel words or the word 'toy', consistent with previous usage in that condition): 'Oh look, there, on the blanket! There's the ______! Can you go get just this ______ for me?' If the child tried to take both objects, E2 or E1 told her to take 'Just the ______.' E2 and the child then played with the chosen object (no feedback was given on whether it was the correct one or not) for a few moments.

After each experimental condition, children participated in a baseline condition. In each of the baseline conditions, E2 and the child began by dropping various familiar toys down a chute or spinning them around on a merry-go-round. E2 then announced that she wanted to drop or spin a different toy now, with the way she did this differing in the different conditions. In the two Knowing baseline conditions, E2 specified exactly which toy she wanted, by asking for the 'frog' (Familiar Word) or the 'toy on the chair' (Designated Generic Word). In the two Guessing baseline conditions, E2 specified that she wanted one toy in particular, specifically 'Now I want to play with the [Bokel]' (Novel Word) or 'Now I want to play with this one toy' (Generic Word). She then looked around the room, turned the child around, and said, 'Oh look, there on the blanket/chair. There's the ______! Can you go get just this ______ for me?' (using the appropriate word). The key point is that in the Guessing baseline conditions, it was impossible for children to know which of the two toys E2 wanted, whereas in the Knowing baseline conditions, it was highly likely that they would know.

For each child in each experimental and baseline condition, there was a different pair of objects, a different container (where appropriate), and a different novel word (where appropriate). Table 3 presents a comparison of the procedures in the different conditions.

Coding and reliability

All trials were videotaped. Children's choices in the retrieval tests for each condition were recorded live on code sheets by E1 and E2 in collaboration, and then checked by an independent observer on the videotapes. Agreement was 100% Videotapes were examined to obtain measures of children's latency to choose an object and their uncertainty about their choice of object in the retrieval tests. The latency (in seconds) to choose an object was defined as the amount of time from when E2 finished her verbal request to when the child gave her final choice to E2 (note that this was a slightly different procedure from Study 1). Uncertainty (yes/no) was coded if children looked questioningly to one of the Es with their choice, asked 'This one?' (or something similar) during their choice, tried to take both objects (with clear uncertainty), or refused to choose. Reliability was assessed for latency and uncertainty by having another coder recode seven randomly selected children (22%) while blind to the experimental condition. For uncertainty, agreement using Cohen's kappa was 0.74. For latency, a Pearson's product-moment correlation between the latencies identified by each coder was 0.98, $p < .001$, and a t-test revealed no significant difference between the two sets of scores ($t = 1.34$, n.s.; the difference in means was 0.14 s).

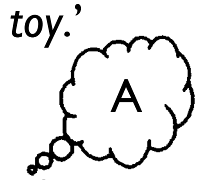
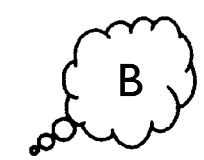
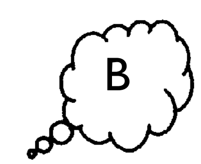
Results

We analysed the effect of the order of presentation of each condition (all four false belief and all four baseline) for the following variables: performance, latency to respond, and signs of uncertainty. Owing to the small samples involved in the performance and uncertainty variables, we collapsed the first and second orders and compared them with the collapsed third and fourth orders for each condition. The order in which children experienced the various conditions had no effect on their performance (Fisher tests: $p > 0.14$ in all cases), latency to respond ($F(9,21) < 2.25, p > 0.05$, in all cases), or signs of uncertainty (Fisher tests: $p > 0.37$ in all cases). Consequently, order is not considered in any further analyses.

Main analyses

Figure 3 presents the percentage of children who were correct in each of the four

Table 3. Summary of the procedural differences across conditions in Study 2

False Belief A Novel Word	E1 puts A in container	E2 leaves	E1: (i) takes A out, (ii) puts B in	E1: (i) removes B from container, (ii) puts A and B across room on the blanket	*[Faudi].'*  E2	E2: (i) tries to open container, (ii) looks around and says: *'Oh look—there on the blanket—there's the …*	*[Faudi]* …	*Can you get it for me?'*
False Belief A Generic Word					*toy.'*  E2		*toy* …	
False Belief B Novel Word		E1: (i) takes A out, (ii) puts B in	E2 leaves	E2: returns, looks to container, and says, *'I'm going to get the* …	*[Bokel].'*  E2		*[Bokel]* …	
False Belief B Generic Word					*toy.'*  E2		*toy* …	

experimental conditions. First, we analysed whether children performed better when the adult used a novel (vs. a generic) word. We thus compared the Novel Word to the Generic Word condition, collapsed across version of the False Belief test (False Belief A vs. False Belief B—children thus had a score of 0–2 in each word condition). There was no significant effect of the type of word ($t = 0.80$, d.f. = 31, n.s.). Next, we compared the two versions of the False Belief test (collapsed across Novel and Generic Word, so children had a score of 0–2 in each of the false belief versions). There was no significant difference between children's scores in the two versions ($t = 0.19$, d.f. = 31, n.s.). More important is the fact that children were above chance in both false belief conditions (recall that it is not a problem to compare children's responding with chance in this study, because when there is no pull of the real, and children are truly guessing, chance should be 50%). In both the conditions, False Belief B (mean = 1.34) and False Belief A (mean = 1.31) children were above the chance value of 1.00 (False Belief B: $t = 2.78$, d.f. = 31, $p = 0.009$; False Belief A: $t = 3.30$, d.f. = 31, $p = 0.002$). This indicates that children were able to choose the last object that E2 had seen regardless of whether it was the first or the last in the container.

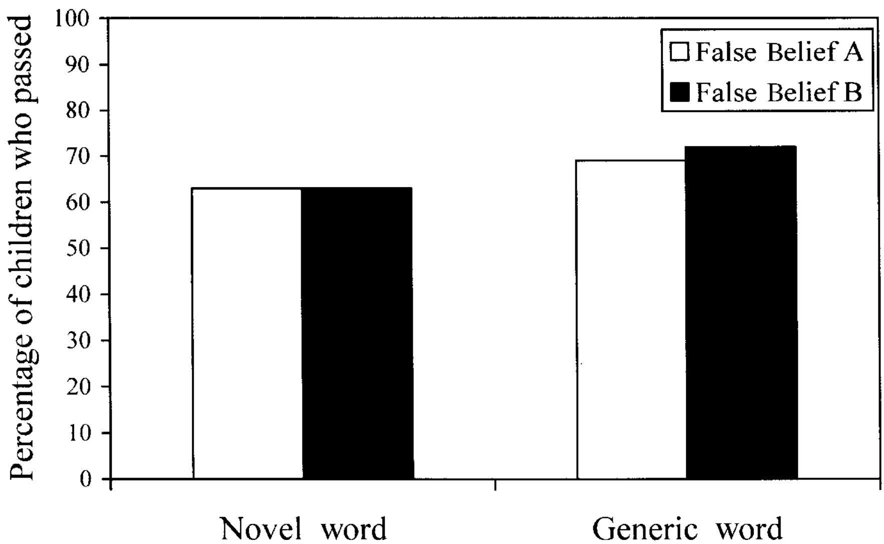


Figure 3. Percentage of children who passed each condition in Study 2.

Finally, we collapsed the data across all conditions, which gave each child a score of 0–4. Children performed significantly above the chance value of 2.00 with this combined score (mean = 2.66; $t = 4.49$, d.f. = 31, $p < 0.001$). Individual patterns indicated that 20 children (63%) responded correctly in either three or four trials whereas only 3 children (9%) showed the reverse pattern (i.e. responded correctly in one or no trials, binomial test: $p < 0.001$). The nine remaining children responded correctly to two trials. Assuming, conservatively, that 3 (9%) of the 20 (63%) children who passed three or four trials did so by chance, we can conclude that 17 (53%) of children were really responding correctly. The majority of children (both as a group and individually) thus passed this test satisfactorily.

Although the current study had a very small age range, we nevertheless checked for any age differences in performance. There were none (Spearman $r = 0.12$, n.s., $N = 32$). Children did not choose the same object during the preference and retrieval tests above chance in any of the four experimental conditions (Binomial tests: n.s.).

Latency and uncertainty

Unlike in Study 1, in this study, we had a way to validate the measures of latency and uncertainty. In the two Knowing baseline conditions, children were faced with the task of retrieving an object the adult asked for with a request they almost certainly understood. In contrast, in the two Guessing baseline conditions, children were faced with the task of retrieving an object that the adult failed to specify adequately. Our prediction was that children would take longer to retrieve the object and show more signs of uncertainty in the Guessing baseline conditions than in the Knowing baseline conditions. This prediction was supported, and thus the measures of latency and uncertainty were validated. For validating both measures, we used only children who correctly retrieved the asked-for object in the Knowing baseline conditions, to make sure that we were dealing only with children who fully understood the adult's request. We collapsed the two versions (Familiar/Novel Word and Generic Word) of each type of baseline condition and used mean scores. The outcome was that children took significantly longer to choose an object in the Guessing condition than in the Knowing condition, $t = 2.17$, d.f. = 17, $p < 0.03$ (see Fig. 4), and they showed more signs of uncertainty in the Guessing condition (mean = 0.29) than in the Knowing condition (mean = 0.13), $t = 2.05$, d.f. = 18, $p < 0.03$ (see Fig. 5; both comparisons one-tailed). Thus, taking a long time to respond, and responding with uncertainty, were associated with guessing.

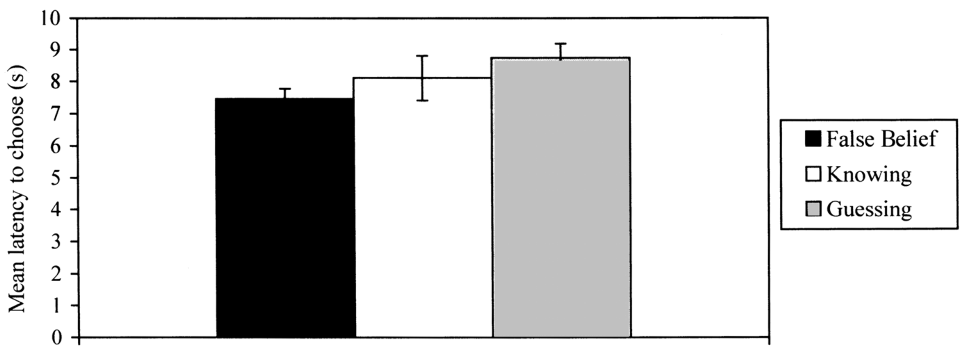


Figure 4. Mean latency to choose in Study 2.

The question of most importance is whether children's latencies and signs of uncertainty in the False Belief conditions looked more like those from the Knowing or the Guessing baseline condition. For latency, the answer is perfectly clear and consistent with our predictions (see Fig. 4). Children chose significantly faster in the False Belief condition than in the Guessing baseline condition ($t = 4.19$, d.f. = 17, $p < 0.001$), and there was no difference between their latencies in the False Belief condition and the Knowing baseline condition ($t = 0.22$, d.f. = 16, n.s.). There was no

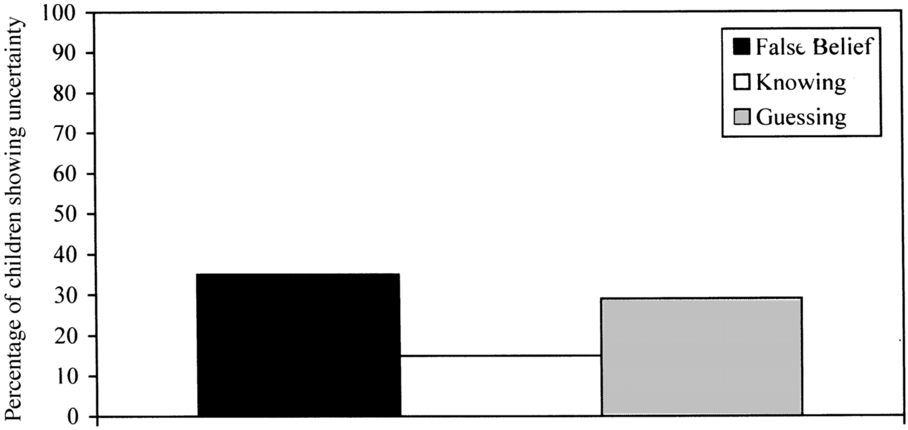


Figure 5. Percentage of children who showed signs of uncertainty in Study 2.

relation between children's False Belief score and their latency to choose (Spearman $r = -0.14$, n.s., $N = 30$; analyses within each of the four conditions produced an analogous result, all t-test: n.s.). Therefore, assuming that children were indeed guessing in the Guessing baseline condition and certain of themselves in the Knowing baseline condition, the pattern of children's latencies strongly suggests that in the False Belief condition, they were not guessing but rather they thought they know which object the adult wanted (even if they were wrong).

The results for uncertainty were slightly different (see Fig. 5). In this case, children in the False Belief condition behaved more like they did in the Guessing than in the Knowing baseline condition. That is, they produced fewer signs of uncertainty in the Knowing baseline than in the False Belief condition ($t = 3.47$, d.f. $= 19$, $p < 0.001$), and there were no significant differences in signs of uncertainty between the Guessing baseline and the False Belief condition ($t = 0.27$, d.f. $= 18$, n.s.). As before, there was no relation between children's False Belief score and their uncertainty score (Spearman $r = -0.14$, N.S., $N = 30$; analyses within each of the four conditions produced an analogous result, all t-tests: n.s.). These results are discrepant with those from the latency measure. It is likely that this is because children's expression of uncertainty has a very low threshold in such difficult tasks.

And so, together, the latency and uncertainty measures tell the following story. In the False Belief condition, children went to make their choice relatively quickly, as quickly as in the Knowing baseline control. But upon actually picking up an object, they checked with the adult as often in the False Belief condition as they did in the Guessing baseline condition. One hypothesis is thus that in the most objective measure of certainty, that is, the speed with which they make their choice, children in our experimental conditions act as if they know what they want—although they are not quite as sure of themselves as they are when the adult explicitly asks for an object with which they are highly familiar.

Comparison between studies

Figure 6 shows the results for the traditional tests and the False Belief conditions from Study 1 and 2. On the traditional Sally-Anne and Smarties tests, children's performance was significantly below 50% chance ($t = 6.86$, d.f. = 28, $p < 0.001$) even when the most lenient criterion for passing was used (i.e. passing only the test question and not necessarily the control questions). In fact, only one child passed both tests using this criterion (7 children passed one test). Children gave correct answers in the two traditional false belief tasks significantly less often than in the False Belief conditions in both Study 1 (binomial tests: $p < 0.001$ both for the strict criterion—test question and all control questions correct—and for the lenient criterion) and Study 2 (for the strict criterion: $t = 12.0$, d.f. = 28, $p < 0.001$; for the lenient criterion: $t < 8.4$, d.f. = 28, $p < 0.001$). An additional piece of information is the relation between the new false belief tests and the traditional tests in Study 2. One-third of children who passed three or four of our false belief tests also passed one or two of the traditional tests, whereas 18%of the children who passed one or two of our tests did this (no children failed all of our tests). The one child who passed both traditional tests passed three of our false belief tests. Thus, children who scored higher on our tests also tended to score higher in the traditional false belief tests, but owing to the small number of children passing the traditional tests, these results did not reach significance (Fisher test: $p = 0.33$, one-tailed).

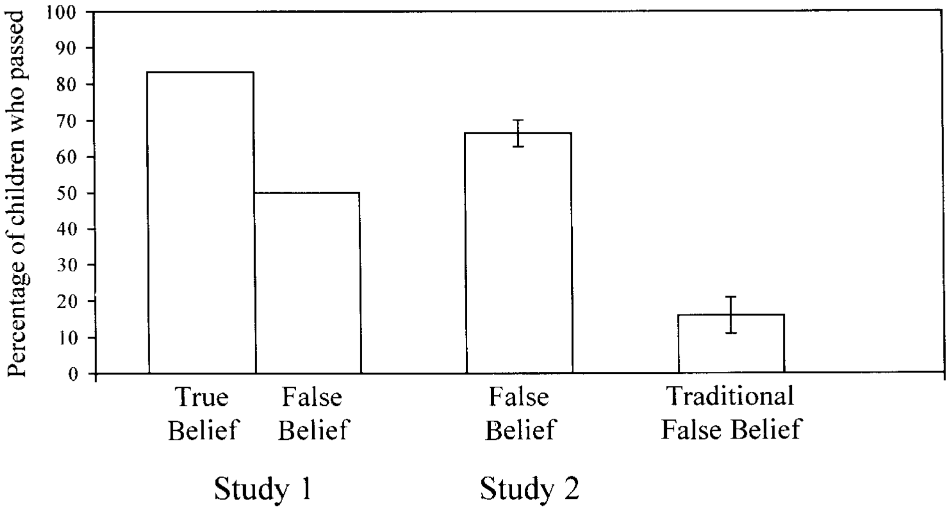


Figure 6. Comparison of children's performance in Study 1, Study 2, and the traditional false belief tests from Study 2. For the Traditional False Belief test, also from the current Study 2, this is the percentage of children who passed the test question only (i.e. whether or not they passed the control questions), averaged across both tests.

Because the methods of Study 1 and Study 2 were highly similar, it is also useful to compare the results of the two studies. Figure 6 also illustrates children's performance in our new True Belief and False Belief conditions from Study 1 and the combined False Belief condition from Study 2. In the False Belief condition of Study 2, children performed better than in the False Belief condition of Study 1 ($t = 4.5$, d.f. = 31, $p < 0.001$) and worse than the True Belief condition of Study 1 ($t = 4.5$, d.f. = 31, $p < 0.001$).

Discussion

The results of this study are clear. As a group, the 36-month-old children performed at higher-than-chance levels on our new false belief tasks. Almost two-thirds of the children responded correctly on all, or almost all, of the four trials, even though this required them to switch strategies and choose a different object (i.e. the first versus the second in the container) on some trials. The children in this study were clearly not exceptional, as their performance on two traditional false belief measures was predictably abysmal. We therefore replicated the main finding from Study 1 that a substantial percentage of 36-month-olds seem to be able to take account of an adult's false belief when attempting to react appropriately to her request for an object.

We also validated the latency and uncertainty measures. Children took longer and showed more signs of uncertainty in the Guessing baseline than in the Knowing baseline. Most importantly, in both false belief conditions, children's latencies patterned with the Knowing baseline, presumably because children were almost as certain of themselves in these experimental conditions as when they knew what the adult wanted for certain. In their expressions of uncertainty, however, the children showed that the experimental conditions were indeed slightly more difficult for them than the as-easy-as-it-gets Knowing baseline. These findings are important for interpreting the current study, but because children's behaviour was so clear (children were above 50%chance in all conditions), they are not crucial. However, this validation procedure has very important consequences for the interpretation of Study 1, where the results were not quite as clear. We will return to this issue in the General Discussion.

We found that using a novel word versus using a familiar, non-specific word made no difference in children's performance. Thus, the strong version of our word-learning hypothesis—that attempting to discern what adults mean when they use novel words puts children in the mindset of 'mindreading'—was not supported. However, in both the Novel Word and Generic Word conditions, children had to determine which object the adult was referring to. Thus, a weaker version of this hypothesis—that determining adults' communicative intentions (whether they involve novel words or uninformative words) puts children in such a mindset—is still a viable possibility.

GENERAL DISCUSSION

In the current two studies, we use a new approach to test young children's understanding of others' false beliefs. Whereas previous studies have used deception, pretence, or looking measures to assess this understanding in 3-year-olds, our tasks are closer in terms of response demands to the traditional tests, and we, too, show that at least some 36-month-old children demonstrate an understanding of false beliefs. Depending on the specific criteria used, we estimate that something over one-third of children understand false beliefs as they are turning 3 years of age. This claim is based on the fact that in both of our studies, a substantial percentage of children not only responded appropriately to object requests that required an understanding of the adult's false belief but also responded correctly (and differently) to object requests in one or another kind of control condition. In addition, a methodological advance of the current study was the use of children's latencies and expressions of uncertainty to disambiguate their performance. A child could have been wrong (or right) on one of our false belief tasks either because they were guessing and made an unlucky (or lucky) guess, or because they had a firm idea about what was or was not the case. Our analysis of children's latencies and expressions of uncertainty in this study indicated that most of the children were not guessing, but rather they were acting quickly and confidently—even if wrong—on their understanding of the other person's belief state.

Some researchers will be sceptical of the findings of Study 1, because as a group, children in the false belief condition did not differ significantly from 50%. However, 50% is not a valid comparison value in this and many other false belief tasks, in which the correct answer conflicts with the current reality as the child knows it (i.e. when there is a pull of the real). Children have a very strong tendency to assume that the adult is seeking the object that is currently in the container, and so when they do something different—when they judge that the adult is seeking the object that is no longer in the container—this judgment must be seen in the light of their pre-existing tendencies. We believe that 3-year-old children as a group are significantly below chance in traditional false belief tasks precisely because of this pre-existing tendency. We thus believe further that children's performance, not only in our study but also in many other studies, should be evaluated not against a chance value of 50%, but rather against the performance that we would expect from them if they did not understand the adult's false belief, that is, their tendency to follow the pull of the real. In this light, there are many studies, including some using traditional tasks, in which it is likely that some 3-year-olds really do understand the adult's false belief.

Others will be sceptical of our interpretation of Study 2, because it does not require children to overcome the pull of the real. Although traditionally many researchers have required children to overcome the pull of the real in order to pass false beliefs tests, it could be argued that children can still demonstrate an understanding of false beliefs in the absence of the pull of the real. To be successful in our study, the child had to determine which of two objects an adult was requesting. To make this determination, the child had to know that the adult believed, falsely, that the last object she saw go into the container was still in there. In fact, in both conditions, the last object the adult saw go into the container was actually across the room (along with the distracter object). Indeed, it is interesting to note that about one-quarter of the children responded spontaneously to E2's initial struggles with the container by attempting to tell her directly that the object she thought was in the container was actually across the room. The main difference of this task from the traditional task is that, in the current version, the real situation as children knew it did not pull them towards choosing the incorrect object as it would have in the traditional task. But why should resisting this pull be the acid test for false belief understanding? The original idea historically was to see whether children understand beliefs, and false belief tasks were used as a way of ensuring that young children were not responding egocentrically (i.e. attributing to the adult the same belief they had; Dennett, 1978). In our study, if the child attributed their own belief to the adult (the container is empty), they would have been totally confused and presumably chosen randomly in the retrieval task.

One other difference between our task and the traditional task is that in our task, children did not need to predict what someone else would do. Some would argue that predicting another's behaviour on the basis of their false belief is an integral part of the false belief task. However, we believe that it is, again, an extra requirement that goes beyond children's false belief understanding. In the current tasks, although children did not need to use their understanding of the adult's false belief to predict what she would do next, they still needed to attribute a belief to the adult that they knew was not true.

There is one possible alternate explanation of our studies that does not involve understanding of others' false beliefs. Children could pass these tasks by simply associating the last object E2 saw with E2 and then assuming that that was the object she wanted. However, it should be noted that this explanation is also possible with the traditional false belief tests (e.g. Flavell, 1988). But recent studies have shown that children use more than a simple seeing-leads-to-knowing rule to pass these tests (e.g. Lohmann, Call & Carpenter, 2001; Garnham & Ruffman, 2001), and so it is unlikely that this is all that children are doing in the current studies.

Although 3-year-olds' difficulties with the traditional false belief task are well known, it is also well known that children of this age routinely display some knowledge of other peoples' mental states in their spontaneous behaviour and discourse (Bartsch & Wellman, 1995; Dunn, 1988). We thus believe that our findings have strong face validity, and that, as many people have objected, the traditional false belief task possesses some irrelevant and very difficult task demands. Our tasks allowed 3-year-olds to display their false belief understanding, because we removed some of these irrelevant task demands. Specifically, in both our studies, children could express their knowledge in the well-practised and well-understood behavioural format of fulfilling an adult's request to retrieve an object. Thus, in our study, there were few of the mismatches of adult and child pragmatic presuppositions and understandings that can often creep into experimental situations in which adults engage children in what for them are unusual types of discourse and conversation (Siegal, 1991). In addition, in most traditional tests, children are asked to predict what a story character will do. But research in a number of different content areas has shown that children have difficulty in using their knowledge of abstract concepts to predict what someone else will do in a hypothetical situation (Mitchell, 1996). Perhaps this is because children of this age have only limited skills in imagining complex situations as they unfold in time, especially if they directly conflict with the current perceptual situation (Harris, 1991, 2000). In our version of the false belief task, we asked children to use their knowledge of another person's false belief in an active behavioural decision, but without explaining or predicting. The first study retained the pull of the real, and nevertheless, more than one-third of the children were able to demonstrate false belief understanding. The second study eliminated the pull of the real, and over half of the children were able to demonstrate false belief understanding.

Some theorists might consider what we have found an 'implicit' version of the real thing, with the real thing being the use of language or gesture to predict what a person with a false belief will do in a hypothetical situation—as in traditional false belief tasks. However, the term 'implicit' is currently used in many different ways and so at the moment is an extremely vague term. It clearly does not mean for most people simply non-verbal, because even in the traditional tasks, children can simply point to where the misinformed person will look, and this knowledge is considered explicit (see Call & Tomasello, 1999, for a non-verbal task that correlates highly with traditional tasks). However, implicit might more plausibly be construed to mean non-behavioural in the sense of 'looking time' measures. Thus, in Clements and Perner's (1994) self-described study of implicit understanding—which found false belief understanding at around the same age as the current study (actually a bit older)—children's expectancies as manifest in their looking times do not reflect any active behavioural decisions; children simply looked where they expected things to happen. Indeed, Ruffman, Garnham, Import, and Connolly (2001) have shown that when younger children look expectantly in these ways, they do not have conscious access to their implicit understanding. In general, there is much current controversy over the interpretation of studies in which children show sensitivity to violated expectations by looking longer at something (e.g. Hood, Carey, & Prasada, 2000), with some researchers believing that it does not reflect genuine knowledge but only perceptual schemas and strategies (Haith, 1998). It may be that being surprised by something is not the same thing as understanding it.

However, in our studies, the child had to make an active behavioural decision based in large part on their understanding of the adult's goal-directed behaviour and communicative intentions as manifest partly in language. The child had to retrieve for the adult the object the adult asked for—when there were two alternatives equally available—and they had to comprehend the adult's linguistic request in order to do so. Given this reliance on language and behavioural decision-making, we believe that the knowledge children demonstrated in our study is not best thought of as implicit. Instead, we see it as falling somewhere in between implicit measures and those that require children explicitly to predict or explain what another person will do on the basis of their false belief.

If our interpretation of our own studies and those of other researchers is correct, a substantial percentage of children at around their third birthdays, and perhaps before, understand false beliefs well enough to use them in their active behavioural decision-making. Other children, however, do not seem to attain this understanding for some time, perhaps 1 or 2 years later in some cases (since many studies find 5-year-old children who still struggle at understanding false beliefs). These facts imply that (1) false belief understanding shows significant individual differences at any given point in the preschool years, and (2) within individuals, the ability to employ knowledge of false beliefs in different discourse settings undergoes significant development during this same period. Although there has been some recent discussion of individual differences in false belief understanding and their possible causes (e.g. Cutting & Dunn, 1999), and other discussions of the sometimes long developmental course of children's coming to understanding others as mental agents (Harris, 1996, 2000), there have not been enough discussions, in our opinion, of the theoretical implications of these facts.

Most directly, if we are talking about a developmental phenomenon whose emergence is both protracted in individual children's development and which emerges in different children at fairly different ages, then modular, maturational accounts of false belief understanding (e.g. Baron-Cohen, 1995) become less plausible. More consistent with these facts are accounts that stress that children's social-cognitive skills emerge from the kinds of complex social and discourse interactions in which they must communicate with interactants of all kinds with all kinds of different background knowledge and expectancies (Appleton & Reddy, 1996; Harris, 1996; Tomasello, 1999). Future research should be aimed at determining with more precision the different developmental courses that false belief understanding can take in different social environments.

Acknowledgements

The authors would like to thank Shannon Tuohy, Brenda Hailey, Carrie Burstein, Solvejg Kuehnert, and Angela Loose for help with data collection; Simone Klein and Susanne Mauritz for help with coding; Jane Childers, Ignae Thomas, and Suzanne Carpenter for help recruiting children; and the teachers and children at the day-care centres for their cooperation.

References

Akhtar, N., & Tomasello, M. (1996). Twenty-four-month-old children learn words for absent objects and actions. *British Journal of Developmental Psychology, 14*, 79–93.

Appleton, M., & Reddy, V. (1996). Teaching three-year-olds to pass false belief tests: A conversational approach. *Social Development, 5*, 275–291.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*, 37–46.

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*, 381–395.

Cassidy, K. W. (1998). Preschoolers' use of desires to solve theory of mind problems in a pretense context. *Developmental Psychology, 34*, 503–511.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language and family background: Individual differences and interrelations. *Child Development, 70*, 853–865.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568–570.

Dunn, J. (1988). *The beginnings of social understanding*. Oxford: Blackwell.

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244–267). Cambridge: Cambridge University Press.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science, 4*, 94–100.

Haith, M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development, 21*, 167–180.

Harris, P. L. (1991). The work of the imagination. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 283–304). Oxford: Basil Blackwell.

Harris, P. (1996). Desires, beliefs, and language. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge: Cambridge University Press.

Harris, P. L. (2000). *The work of the imagination*. Malden, MA: Blackwell.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development, 71*, 1540–1554.

Koos, O., Gergeley, G., Csibra, G., & Biro, S. (1997, April). *Why eating Smarties makes you smart? Understanding false belief at the age of 3*. Poster presented at the meetings of the Society for Research in Child Development, Washington, DC.

Lohmann, H., Call, J. & Carpenter, M. (2001). Seeing versus believing in false belief tasks. *Poster presented at the meeting of the Society for Research in Child Development, Minneapolis, MN, April.*

Mitchell, P. (1996). *Acquiring a conception of mind: A review of psychological research and theory*. Hove, East Sussex: Psychology Press.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Development, 60*, 689–700.

Polak, A., & Harris, P. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*, 561–568.

Robinson, E. J., & Mitchell, P. (1992). Children's interpretation of messages from a speaker with a false belief. *Child Development, 63*, 639–652.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology 80*, 201–224.

Siegal, M. (1991). *Knowing children: Experiments in conversation and cognition*. London: Psychology Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Received 16 March 2000; revised version received 26 October 2001